NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

LEONOR PINHÃO ALMEIDA

BSc of Science in Biomedical Engineering

ELECTROCORTICOGRAPHY SIGNAL SYNTHESIS FOR ENHANCED INTRA-OPERATIVE DETECTION OF EPILEPTIFORM ACTIVITY

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
September, 2024

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

ELECTROCORTICOGRAPHY SIGNAL SYNTHESIS FOR ENHANCED INTRA-OPERATIVE DETECTION OF EPILEPTIFORM ACTIVITY

LEONOR PINHÃO ALMEIDA

BSc of Science in Biomedical Engineering

Adviser: Hugo Filipe Silveira Gamboa
Full Professor, NOVA University of Lisbon

Co-adviser: Luís Miguel Domingues Ferreira Silva
Postdoctoral Researcher, NOVA University of Lisbon

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
September, 2024

Electrocorticography Signal Synthesis for Enhanced Intra-Operative Detection of Epileptiform Activity

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v7.1.29) [1].

À minha família,

Acknowledgements

First, I would like to express my gratitude to Professor Hugo Gamboa and Dr. Luís Silva. To Professor Hugo, for offering me the opportunity to be a part of this project and for believing in me from the very beginning. To Luís, for his unwavering support, presence, and mentorship. He has continually encouraging me to grow and always believed in my potential, guiding me through this challenge and teaching me the value of hard work and resilience — all with a smile. To Inês Silveira, for her constant support and wisdom, for always advocating for me and encouraging me during moments of self-doubt. I owe a special thanks to Dânia Furk, whose unwavering help and kindness have been invaluable — words cannot fully express my gratitude, a simple thank you is not enough. To Sara and Márcia, who shared this journey with and made the toughest times easier, their help and reassurance were truly appreciated. And to all the wonderful people I have met in the LIBPhys Lab, who made my days a little brighter — Phillip, Mariana, Rodrigo, and Jéssica.

On a more personal note, I would like to express my heartfelt gratitude to all my friends. To Ana, for her kindness and love, for always providing me comfort and clarity when I needed it most. To Valentina, for consistently reminding me of the bright side of life. To Margarida, for being like a sister to me and always making me feel at home — this would not have been possible without you. To my island friends, for their constant support, even from afar. To Rita, my best friend, roommate, and confidant over the past five years, I am forever thankful to have walked this path side by side with you. Finally, to Guilherme, for his love and support, and for always inspiring me to strive to become the best version of myself.

Above all, my deepest gratitude goes to my family, whose love and support were always felt, even across the ocean. To my mom, who has always gone above and beyond to ensure her daughter lives a life filled with happiness, love and laughter. To my dad, who taught me to never doubt myself and whose constant belief in me has motivated me every step of the way. And to my brother, for always taking care of me and keeping me in check throughout these years. I truly appreciate everything you have done for me. Everything I am, I owe to the three of you.

Abstract

Epilepsy surgery is often necessary in drug-resistant cases where anti-seizure medications are ineffective. A key challenge during surgery is accurately localizing the epileptogenic tissue to be resected, typically done through visual assessment of epileptiform activity in intra-operative electrocorticogram (ioECoG) signals. While Machine Learning (ML) holds promise for automating this detection, data scarcity limits the performance of these models. To overcome this limitation, this thesis introduces a generative model to synthesize ioECoG signals, aimed to produce realistic representations of both epileptic and healthy brain tissues, ultimately advancing the development of reliable tissue classification systems.

To achieve this, three different data patterns were distinguished with Agglomerative Clustering, which were then fed into individual generative models based on a Deep Convolutional Wasserstein Generative Adversarial Network with Gradient Penalty (DCwGAN-GP). The synthetic data was then evaluated across three domains: Fidelity, Diversity, and Utility. In terms of Fidelity, the synthetic signals successfully captured the temporal and frequency characteristics of the real data, replicating common ioECoG patterns. However, Diversity analysis using Principal Components Analysis (PCA) and T-distributed Stochastic Neighbour Embedding (t-SNE) revealed that the synthetic data lacked variability compared to the real dataset, likely due to misclassified real data that the generative model did not capture. Finally, the Utility of the synthetic data was demonstrated by training several classifiers on both real and synthetic data to distinguish between the three previously identified clusters. Classifiers trained on the combined dataset achieved accuracy scores of 96%, compared to 89% when trained solely on real data.

These results highlight the relevance in using synthetic data to augment existing ioECoG datasets and enhance the performance of ML classifiers. This advancement enables a more accurate identification of the epileptogenic zone, leading to improved surgical outcomes and a better quality of life for the patients suffering from this disease.

Keywords: Synthetic Data, Epilepsy, Epileptiform Activity, Epileptogenic Tissue, ioECoG, GAN

Resumo

O tratamento cirúrgico da epilepsia torna-se necessário quando os medicamentos antiepiléticos não são eficazes em casos de epilepsia refratária. Um dos maiores desafios durante esta cirurgia é a localização precisa do tecido epiléptico a ser removido, normalmente realizada por meio da inspeção visual dos sinais de eletrocorticografia intraoperatória (ioECoG). Embora os métodos de aprendizagem automática apresentem grande potencial para automatizar essa localização, a escassez de dados limita a eficácia desses modelos. De modo a superar esta limitação, esta dissertação propõe um modelo generativo capaz de sintetizar sinais de ioECoG, com o objetivo de produzir representações realistas de tecido cerebral saudável e epiléptico, promovendo o desenvolvimento de sistemas de classificação mais precisos.

Para atingir este objetivo, foram identificados nos dados três padrões distintos com a utilização do algoritmo de *Agglomerative Clustering*. Esses padrões foram incorporados em modelos generativos individuais, com o uso de uma *Deep Convolutional Wasserstein Generative Adversarial Network with Gradient Penalty* (DCwGAN-GP). Os dados sintéticos gerados foram posteriormente avaliados em três dimensões: Fidelidade, Diversidade e Utilidade. No domínio da Fidelidade, os dados gerados capturaram as características temporais e de frequência dos sinais reais. Contudo, a análise da Diversidade, realizada através da Análise de Componentes Principais e da Distribuição Estocástica em Vizinhança t-Distribuída, revelou uma menor diversidade, possivelmente devido amostras reais mal classificadas. Por fim, a Utilidade foi demonstrada ao treinar classificadores com dados reais e sintéticos, onde os resultados mostraram uma exatiadão de 96%, comparada com os 89% obtidos ao utilizar apenas dados reais.

Estes resultados demonstram que o uso de dados sintéticos possui a possibilidade de melhorar a localização do tecido epiléptico, contribuindo para melhores resultados cirúrgicos e consequemente uma melhor qualidade de vida dos pacientes.

Palavras-chave: Epilepsia, Dados Sintéticos, Atividade Epileptiforme, Tecido Epileptico, GAN

Contents

List of Figures

List of Tables

Acronyms

1D one-dimensional *(pp. 10, 18, 19)*

ASMs anti-seizure medications *(pp. 1, 3)*

cGAN Conditional Generative Adversarial Network *(pp. 9–11)*
CNN Convolutional Neural Network *(pp. 18, 38)*

DCGAN Deep Convolutional Generative Adversarial Network *(pp. 9, 10)*
DCwGAN-GP Deep Convolutional Wasserstein Generative Adversarial Network with Gradient Penalty *(pp. 11, 17, 19, 20, 39, 40)*
DL Deep Learning *(pp. 5, 11)*

ECoG electrocorticography *(pp. 4, 34)*
EEG electroencephalography *(pp. 1, 3–5, 8–11, 14, 15, 18, 28, 35)*
ES epileptic seizures *(pp. 3, 9, 10)*
EZ epileptogenic zone *(pp. 3, 4, 8)*

FFT Fast Fourier Transfrom *(pp. 14, 16, 25, 26)*
FID Fréchet Inception Distance *(pp. 10, 11)*

GAN Generative Adversarial Networks *(pp. 5, 6, 9, 11, 39, 40)*

HFO High Frequency Oscillations *(pp. 4, 5, 14, 15, 36)*

ioECoG intra-operative electrocorticogram *(pp. iv, viii, 1, 2, 4, 5, 8, 11–15, 18, 21, 25, 34–36, 38–40)*

LSTM Long Short-Term Memory *(pp. 10, 11, 39)*

ML Machine Learning *(pp. iv, 1, 2, 5, 7–9, 15, 23, 40)*
MRI magnetic resonance imaging *(pp. 1, 3)*
MSE Mean Squared Error *(p. 37)*

NN Neural Networks *(pp. 5, 8, 23, 24, 32)*

PCA Principal Components Analysis *(pp. iv, 22, 31, 32, 37)*
PET Positron Emission Tomography *(pp. 1, 3)*
PSD Power Spectrum Density *(pp. 9, 10, 14, 16, 21, 26–28, 30, 35, 37, 38, 40)*

RD Real Data *(p. 12)*
RF Random Forest *(pp. 11, 24, 32, 33)*
RR Real-Real *(pp. 28, 30, 31, 36, 37)*
RS Real-Synthetic *(pp. 28, 30–32, 36, 37)*

SD Synthetic Data *(pp. 2, 12)*
SEEG stereoelectroencephalography *(pp. 4, 10)*
SS Synthetic-Synthetic *(pp. 28, 30, 31, 36, 37)*
SSIM Structural Similarity Index *(pp. 30, 31, 37)*
STD standard deviation *(pp. 21, 27, 28, 30, 35–38)*
SVM Support Vector Machines *(pp. 23, 24, 32)*

t-SNE T-distributed Stochastic Neighbour Embedding *(pp. iv, 11, 22, 31, 32, 38)*

VAE Variational Auto-Encoders *(pp. 9–11)*

WCSS Within-Cluster Sum of Squares *(pp. 16, 25)*
WD Wasserstein Distance *(pp. 6, 10, 21, 28, 30, 36, 38)*
wGAN Wasserstein Generative Adversarial Network *(pp. 6, 10, 11)*
wGAN-GP Wasserstein Generative Adversarial Network with Gradient Penalty *(pp. 9–11)*

1

Introduction

1.1 Context and Motivation

Epilepsy is one of the most common neurological diseases, affecting approximately 50 million people globally [2]. Although anti-seizure medications (ASMs) are the primary treatment to be considered for epilepsy [3], about a third of patients do not achieve seizure control through medication, resulting in drug-resistant epilepsy [4]. For these patients, epilepsy surgery offers a viable solution, potentially curing seizures and enhancing their quality of life [3].

The key concept of this surgery is the identification and removal of the area of the brain that generates seizures [4, 5]. This can be achieved by localizing the affected tissue using intra-operative electrocorticography that enables the distinction of pathological electrographic activity, including spikes, sharp waves and ictiform spike patterns, thereby identifying the epileptic brain tissue [6][7]. The success of the surgery is highly dependent on the accurate localization of this tissue [8]. Despite exhaustive pre-surgical evaluation to properly define this localization [9], including magnetic resonance imaging (MRI), video electroencephalography (EEG), and, if needed, Positron Emission Tomography (PET) or ictal Single-Photon Emission Computerized Tomography (SPECT) [4, 8], labelling brain tissue presents many difficulties. These challenges come from the difficulty of distinguishing epileptiform background patterns from normal ioECoG activity [10], usually performed by clinicians through visual scanning of the signals during surgery [11]. Manual interpretation of these signals is prone to human error [12] and often overlooks valuable information [13], emphasizing the need for a more reliable technique to detect epileptiform activity in ioECoG signals.

In recent years, research about the usage of ML algorithms to evaluate ioECoG signals in order to improve the automatic detection of epileptiform discharges has shown promising results [13] [11] [14] [15] [16] [17]. Nevertheless, gathering sufficient usable ioECoG data is challenging [18] due to: cost expenses (expensive hardware and costly medical expertise), large amounts of participant time and practical challenges regarding the collection itself (infections) [19]. This scarcity of ioECoG data compromises the robustness

and generalisability of ML models, since they require a large amount of data to be able to distinguish pathological patterns from ioECoG background activity, which exhibits high inter-subject variability [20]. In order to overcome these challenges and, therefore, increase the robustness of these ML detection algorithms, the generation of Synthetic Data (SD) emerges as a potential solution [21] [22] [23] [24].

1.2 Objectives

In the context of the Synthetic and Scalable Data Platform for Medical Empowered AI (AISym4Med) project in collaboration with the University of Utrecht, the aim of this study is to develop an ioECoG synthetic data generation system capable of producing signals from epileptic and healthy brain tissue in the context of epilepsy surgery.

More specifically, this work aims to:

- Identify common patterns that might exist in the ioECoG data, providing new insights into epileptiform background patterns;
- Investigate the use of generative algorithms to augment ioECoG data, generating synthetic ioECoG data that resembles epileptic and non-epiletic brain tissue patterns;
- Evaluate the generated data using similarity quality measures;
- Compare the performance of ML models trained with and without the generated data to evaluate its impact on the accuracy and reliability of the models in distinguishing epileptic from healthy tissue;

The research work described in this dissertation was carried out in accordance with the norms established in the ethics code of Universidade Nova de Lisboa. The work described and the material presented in this dissertation, with the exceptions clearly indicated, constitute original work carried out by the author.

2

Theorical Concepts

2.1 Epilepsy

Epilepsy is a central nervous system disorder [25] characterized by an abnormal brain activity in different sites of the brain [2], resulting in recurrent epileptic seizures (ES), atypical behaviours and sometimes loss of awareness [2] [26]. Inherently, these seizures are unpredictable and sporadic [9] and are identified by short episodes of involuntary movements [2]. Diagnosis involves clinical assessment, supported by a series of tests and evaluations to fundament the decision-making process [9]. The primary goal in the treatment is to achieve complete seizure control [27], that is remaining seizure-free for at least one year or having significant reduction in seizure frequency [28]. To achieve this, first-line interventions include ASMs [3]. However, when seizures persist despite the usage of at least two appropriately chosen ASMs, epilepsy is classified as refractory or drug-resistant [9]. Considering the significant deterioration in the quality of life for patients suffering from drug-resistant epilepsy [3], alternative treatments such as surgery, neurostimulation, laser interstitial thermal therapy, and stereotactic radiosurgery have been proposed [9].

2.1.1 Epilepsy Surgery

In individuals with drug-resistant epilepsy, neurosurgical intervention offers a potentially curative approach, aiming to achieve complete seizure control without causing neurological complications [3] [4] [10]. The basic concept in epilepsy surgery is the removal of the epileptogenic zone (EZ), the injured brain tissue triggering seizures [5]. Thus, the achievement of seizure freedom is highly dependent on the accurate localization of the EZ and its complete and safe removal [3]. In order to properly identify this tissue, patients with drug-resistant epilepsy require pre-surgical evaluation, normally done by analysing clinical features, seizure semiology and data from various exams, such as EEG, MRI and PET [4] [9].

2.1.2 Electrocorticography

Electroencephalography is a technique for measuring the brain's electrical activity [29]. While extra-cranial EEG analyses brain activity throughout both brain hemispheres in the scalp, intra-cranial EEG (iEEG) involves targeted recordings directly from surgically implanted electrodes placed in specific brain regions [29]. iEEG can be measured with two different recording methods: stereoelectroencephalography (SEEG) and electrocorticography (ECoG) [30]. In particular, ECoG involves placing electrodes directly over surgically exposed cortical surfaces to precisely identify the EZ for epilepsy surgery. This monitoring can be either performed in the operating room during surgery, termed ioECoG, or it can be done outside the surgery room for a longer time [5]. ioECoG is usually done during epilepsy and tumour surgery to optimize the definition of the EZ through the identification of interictal epileptiform activity, such as interictal spikes or High Frequency Oscillations (HFO) [8].

2.1.2.1 Utilization during Epilepsy Surgery

The use of ioECoG to define the tissue to be resected involves three sessions [8], depicted in Figure 2.1. The first is pre-resection, comprised of the recordings conducted before the planned resection begins, employing various grid and strip/depth configurations. Then, the intermediate session involves recordings performed before any iterative extension of the resection area. Finally, post-resection sessions include recordings after the last resection to confirm the seizure foci and to assess whether the extent of the resection is sufficient [5].

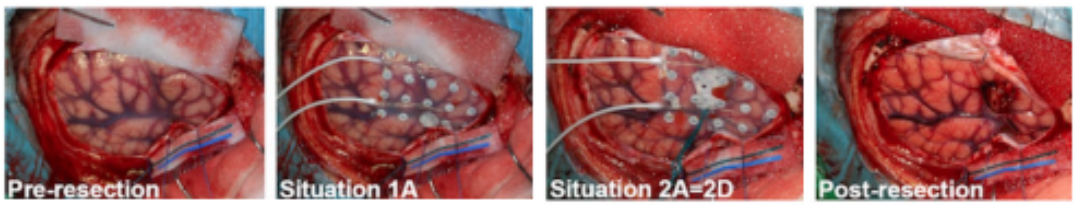

Figure 2.1: Recording Sessions in Epilepsy Surgery with ioECoG. The pre-resection stage of epilepsy surgery is shown in the images labeled Pre-resection and Situation 1A. Situation 2A–2D represents the intermediate session, and Post-resection shows the final stage after the epileptogenic tissue was resected and the grid removed. Adapted from [8].

2.1.2.2 Epileptiform activity

Epileptiform activity includes spikes, sharp and slow waves, and spike-and-slow wave complexes (see Figure 2.2) that are distinct during the ictal, preictal and interictal periods [11] and are strongly associated with epilepsy [31]. With the support of other non-invasive techniques, these patterns have been widely utilized to help identify the EZ and aid in defining the extent of surgical resections [5]. Spikes are defined as sudden sharp transients in the conventional EEG frequency range, between 0 and 80 Hertz, with a

maximum duration of 80 milliseconds and an amplitude of at least twice the baseline [10]. During recent years, it has been shown that estimating the power in the EEG frequency bands can help the distinction between epileptiform and non-epileptiform iEEG signals [12]. In addition, it has been revealed that ioECoG oscillations higher than 80 Hertz contain discriminating information as well, the so-called high frequency oscillations [32]. These HFO spikes represent transient neural activity, classifiable as ripples (80-250 Hertz) or fast ripples (250-500 Hertz) [10], lasting tens of milliseconds [14].

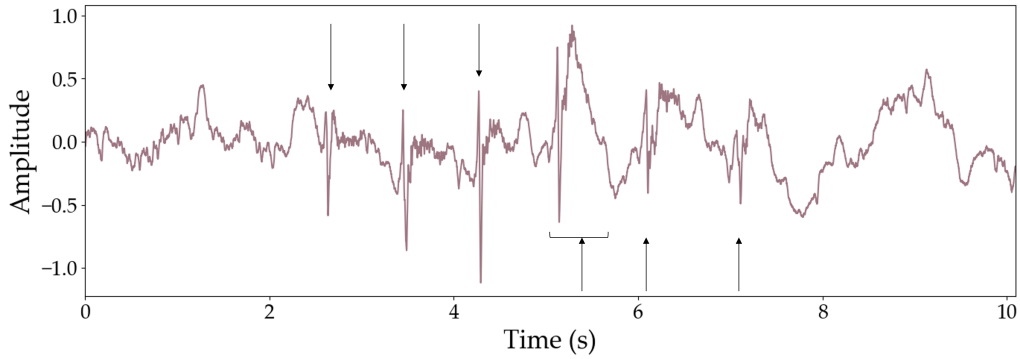

Figure 2.2: Example Waveform of Epileptiform Signals, illustrating spikes and spike-and-slow-wave complexes. Data taken from the study dataset (see Section 4.2.1) and normalized, therefore the amplitude is shown in arbitrary units.

2.2 Synthetic Data Generation

Generative models will be employed to produce synthetic ioECoG data for augmenting ML classification algorithms. Thus, this section will define synthetic data generation using these models, specifically focusing on Neural Networks (NN).

Deep Learning (DL) is a branch of ML that uses NN [33] to uncover complex patterns in large datasets, by using backpropagation to adjust the model's internal parameters, transforming each layer's representation based on the previous layer's output [34]. In turn, synthetic data generation utilizes DL to create artificial data through generative models that resemble real-life observations by learning its distribution [35]. Depending on the research objective, synthetic data can either replace, augment or act as a proxy for real data [36], while maintaining or even improving its utility [35].

2.2.1 Generative Adversarial Networks

Generative Adversarial Networks (GAN) represent a class of deep generative models for unsupervised learning tasks [36], employing two NN: a generator, G, and a discriminator, D [37] [19]. Engaged in a min-max competition [37], the generator attempts to create realistic synthetic samples designed to match the training data distribution [38], while the discriminator aims to distinguish between the real and generated samples [22]. A simplified overview of the GAN architecture can be seen in Figure 2.3 .

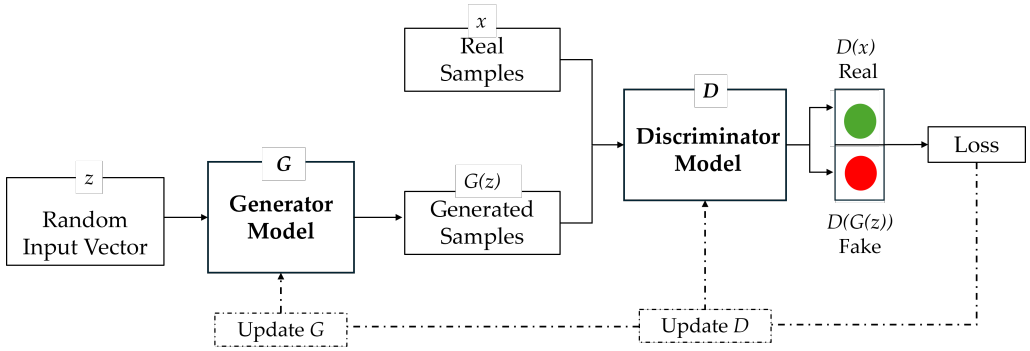

Figure 2.3: Overview of a Generative Adversarial Network Architecture. The generator G creates synthetic samples from a random input z, while the discriminator D distinguishes between real x and generated samples $G(z)$, providing a probability that an input is real, typically within the range of $[0, 1]$ or $[-1, 1]$. As D learns to classify samples, its gradients update both networks through the loss function [39].

This competition can be described with the value function 2.1 [40]. While D is trained to maximize its ability to correctly classify real and generated samples, G is optimized to minimize the function $\log(1 - D(G(z)))$, aiming to produce samples that deceive the discriminator.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]. \tag{2.1}$$

In the end, the goal of GAN is to reach a Nash equilibrium (i.e. a condition where no player can improve their outcome by independently altering their strategy [41]), where the generator captured the precise distribution of real data, creating such realistic samples that the discriminator cannot distinguish whether they are real or synthetic [35] [42] [43].

Despite of GAN's high performance in generating synthetic data, it faces limitations such as training instability, mode collapse [37], and vanishing gradients [22] [24]. In turn, the Wasserstein Generative Adversarial Network (wGAN) provides a solution to these issues, utilizing the Wasserstein Distance (WD) as loss function to minimize the difference between the generated and real data distributions [24]. The cost function of the wGAN can be computed with equation 2.2, where similarly to equation 2.1, the discriminator, now referred as critic, C, seeks to maximize the average cost $E(C(x))$ and G seeks to minimize the average cost $E(G(z))$. As C learns how to classify the data, it returns smaller WD values for real data, and larger values for fake [22].

$$\min_G \max_C \mathbb{E}_{x \sim p_g}[C(G(z))] - \mathbb{E}_{x \sim p_{\text{data}}(x)}[C(x)] \tag{2.2}$$

By measuring the distance between the two distributions as the loss function, the critic's output no longer needs to be constrained to the range $[-1, 1]$, helping to reduce vanishing gradients and mode collapse [22]. To improve training stability and convergence [37], normally the wGAN is paired with a Gradient-Penalty that enforces a *Lipschitz* constraint on the critic's weights, ensuring continuity and differentiability, thus leading to more stable outcomes [22]. The loss function of the wGAN with Gradient Penalty can be

computed by adding the penalty to the original loss function in 2.2 [19], where in 2.3 the penalty term is scaled by λ, the regularization factor.

$$L = \mathbb{E}_{x\sim p_g}\left[C(G(z))\right] - \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[C(x)\right] + \lambda\mathbb{E}_{x\sim p_{\text{data}}(x)}\left[(\|\nabla C(\hat{x})\|_2 - 1)^2\right] \tag{2.3}$$

2.2.2 Evaluation Metrics

After using the generative model to synthesize data, it is paramount to evaluate the synthetic signals in comparison to the real data. Therefore, the evaluation of synthetic data can revolve around three key domains: Fidelity, Diversity and Utility [44] [45] [46]. Together, these domains assess the extent to which the synthetic data replicates real world characteristics, captures a broad range of variations, accurately representing the original data, and serves practical applications effectively.

Firstly, Fidelity measures the similarity between synthetic and real samples, ensuring that the synthetic data accurately reflects the characteristics of the real data [46] [44]. It is essential for evaluating how accurately the synthetic data replicates the details of the real data, without simply duplicating it.

Diversity refers to the capability of the synthetic data to generate a wide range of high-quality samples, encompassing the full variability found in real-world data [45] [46]. A high diversity score indicates that the synthetic dataset can mimic the variation seen in actual datasets.

Lastly, utility evaluates the practical applicability of synthetic data by testing whether a ML model trained with additional synthetic data can outperform the same model trained only on the original data [23]. This domain is essential for assessing the overall utility of the synthetic dataset in real-world ML tasks.

3

Literature Review

The use of ioECoG serves as a tool during epilepsy surgery to assist in the removal of injured tissue in patients with drug-resistant epilepsy. During surgery, clinicians visually analyze epileptiform signals captured by the ioECoG to assess and refine the boundaries of the resection area [47]. However, interpreting these signals is challenging, prone to human error, and time-consuming [15]. This obstacles rise the need for more precise methods to distinguish between epileptogenic and non-epileptogenic tissue [47], leading to the development of signal biomarkers and automatic approaches using ML. Nonetheless, this automatic approaches require a large amount of high-quality training data to ensure their robustness [20]. Yet, collecting high-quality EEG data remains difficult [21] due to the costly, time-consuming processes and the need for complex experimental setups [48] [37]. EEG data also exhibit high variability with physiological meaning and patterns that are subject- and session-dependent [39]. Moreover, due to its invasive nature and potential risk of infection, intra-cranial EEG is only reserved for a reduced number of patients [49], aggravating the scarcity of EEG data. This lack of sufficient data poses significant challenges in training robust ML models [22], resulting in overfitting [20]. In order to address this limitation, various studies have proposed the generation of synthetic EEG data. Therefore, this chapter will explore the implementation of automatic EZ localization approaches and the generation of synthetic EEG data as solutions to the challenges of interpreting ioECoG signals and the scarcity of data.

3.1 Automatic Detection of Epileptic Tissue

Given the challenges in precisely localizing the EZ, several studies have proposed different methods for its localization [13] [11] [14] [15]. Their motivations ranged from the need to classify focal and non-focal EEG signals [13], to automate the detection of epileptiform activity [11] [15] and enhance epileptic tissue localization [14]. To achieve these goals, researchers have employed various approaches, including: a Wavelet Synchronizing Transform-Deep Convolutional NN [13], quadratic classifiers to separate different types of signals [11], skewness-based functional connectivity (SFC) analysis in the high-frequency

band [14] between epileptic and non-epileptic tissue and Random Forest classifiers for spike detection [15]. Pre-processing methods involved techniques such as the application of bandpass (0.5–150 Hz) [13] and low-pass (0–60 Hz) filters, signal normalization [11], bad channel removal, re-referencing, segmentation, and envelope extraction [14]. In turn, feature extraction focused on time, frequency, time-frequency, statistical, and nonlinear features [13], [11] [15]. This included the use of specific frequency sub-bands, dimensionality reduction [11], and time-frequency features such as Hilbert instantaneous amplitude, Fourier-transformed spectral characteristics, and wavelet-based features [15]. These methodologies yielded promising results, with some models achieving accuracy's from 78.8% [15] to 99.7% [13] in the distiction between different epileptic patterns.

3.2 Generation of Synthetic Electroencephalography Data

The generation of high-quality synthetic data for medical applications has become increasingly significant, not only for addressing patient privacy concerns [42], but also due to its potential in augmenting training datasets and enhancing the performance of ML models in various EEG-based tasks [42] [21] [22] [23] [24] [50] [51]. Liang et al. [18] highlighted that common data augmentation methods include GAN, Variational Auto-Encoders (VAE), Diffusion Models, cropping and noise addition. Despite the notable progress made with GAN in synthetic data generation [42], its application in the medical domain, particularly in augmenting EEG data, faces challenges such as mode collapse, complex training processes, and instability [52], [18]. Nonetheless, GAN remains the leading generative model for synthesizing artificial data [18] [23] [53] [52]. Therefore, this section provides a brief overview of the key studies conducted in this area.

Over the past decade, various GAN architectures have been proposed, each attempting to overcome the limitations of earlier models. The conditional variation, Conditional Generative Adversarial Network (cGAN), introduced in 2014 [54], and proven to be effective for training stability and mitigating mode collapse [37], has been commonly utilized in the literature. Pascual et al. [42] utilized this conditional variation to synthesize four second EEG signals for ES prediction while Carrle et al. [23] used it for the diagnosis of depression. Employing Deep Convolutional Generative Adversarial Network (DCGAN), Aznan et al. [21] generated three second dry-EEG signals for a Steady State Visually Evoked Potential task and Rasheed et al. [55] synthesized 60 seconds intracranial-EEG spectrograms for the prediction of ES. The variation of Wasserstein Generative Adversarial Network with Gradient Penalty (wGAN-GP), being the most popular utilized architecture [50] [51], introduced a more stable gradient, leading to more robust training processes and a higher-quality synthetic data [23]. Park et al. [22] utilized this approach to synthesize EEG signals sampled at 160 Hz for motor imagery classification. The model was trained during 50 000 epochs and the signals evaluated through the visualization of their Power Spectrum Density (PSD) and the comparison between the performance of different classifiers. While the study reported improvements in classifier performance with the augmented dataset,

it was also reported that the synthetic signals' PSD lacked fidelity. Recent innovations have sought to further advance these variations. Wang et al. [53] introduced a dual discriminator architecture for synthesizing sleep EEG signals, one focusing on the time-series data and the other on the spectral domain, capturing the information in frequency domain, a critical aspect of sleep EEG analysis. Similarly, Lalli et al. [56] utilized a dual-encoder based cGAN, where it combined the discriminator analysis of the EEG time-series with the VAE reconstruction error. Moreover, Du et al. [57] utilized the variation wGAN-GP incorporating Long Short-Term Memory (LSTM) in the generator network to capture the temporal characteristics present in the time-series. Pascual et al. [48] and Pan et al. [58] utilized a U-Net generator in their architectures, where the latter also utilized a divergence regularization term to regularize the training process and reduce calibration time. Auxiliary Decoding techniques have also been employed to enhance signal quality, for instance Liang et al. [18] and Aznan et al. [39] integrated an auxiliary classification within the discriminator to classify the input class while distinguishing between real and synthetic signals.

More specifically, in 2018, Hartmann et al. [59] introduced EEG-GAN, the first framework for synthesizing EEG signals. This pioneering approach employed an improved wGAN-GP to generate single EEG channels with a sampling frequency of 256 Hz. The generator and critic networks both featured blocks of two convolutional layers, with the generator incorporating an upsampling layer and the critic a downsampling layer. To minimize frequency artifacts, average pooling and strided convolution with a factor of two were implemented. The generator was trained once for every five critic iterations, over 2000 epochs, with an additional 2000 for fading. The synthetic signals were then evaluated using three metrics: Fréchet Inception Distance (FID), Inception Score, and sliced WD, an aproximation of the traditional WD but utilizing one-dimensional (1D) projections. Interestingly, better FID and Inception Score results did not consistently correlate with improved signal quality. However, better sliced WD scores were associated with more realistic generated signals, particularly in terms of their spatial and spectral properties.

Xu et al. [26] employed wGAN, DCGAN and DCWGAN to generate multichannels to enhance the prediction of ES. Here the CHB-MIT EEG dataset, sampled at 256 Hz, was used with signals segmented into 20 second windows and downsampled to 100 Hz. The networks were trained similarly to the study of Hartmann et al. [59], with the generator trained once per five iterations of the critic, during 30000 training epochs. Notably, DCWGAN outperformed the others models across all evaluation metrics: frequency domain root mean square error (FDRMSE), FID, WD and classification performance, generating higher-quality data.

Recently, Wu et al. [24] introduced a novel approach for synthesizing SEEG epileptic signals, employing a conditional transformer-based wGAN-GP. The four-second signals were downsampled to 1000 Hz, and a fourth-order bandpass filter was applied between 0.5 Hz and 400 Hz, along with a notch filter to eliminate 50 Hz powerline noise. The generative network incorporated transformers in both the generator and discriminator to

capture the temporal dependencies of the time series, with a target label in the generator facilitating conditional generation. An additional classification error term was integrated into the Wasserstein loss function, to ensure the generation of distinguishable features across different classes. The synthetic signals were then evaluated using visual inspection with t-SNE, Cosine Similarity (CS), Jensen-Shannon Distance (JSD), and the performance of a DL model to classify five hand movements. The author compared the proposed cTGAN with three traditional augmentation methods: noise injection, VAE, and wGAN-GP. Visual inspection revealed that the first two methods produced inferior samples, solidifying the superiority of GAN models in this context. Furthermore, the proposed method outperformed the wGAN-GP in terms of CS, JSD, and classifier performance.

Du et al. [20] proposed a DCwGAN-GP to synthesize time-frequency representations from EEG signals sampled at 250 Hz. The generator comprised of transposed convolutional layers, while the discriminator consisted in convolutional layers, both with layer normalization and trained throughout 2000 iterations. The synthesized spectograms were evaluated visually and through the performance of two CNN classifiers. While the generated samples were reported to visually exhibit contrasts and contours similar to the real data, they were also noted to have irregular textures and lack naturalness. Despite these results, the combined dataset (synthetic and real) improved the classification performance.

Cook et al. [60] proposed the first architecture for generating EEG signals to predict brain age. The study compared three different models to synthesize signals resampled at 128 Hz and filtered between 0.5 to 55 Hz: cGAN, wGAN and wGAN-GP. The synthetic signals were evaluated through a Kolmogorov-Smirnov test and through the performance of three classifiers to enhance the accuracy of age prediction: Support Vector Regression, LSTM and Random Forest (RF). All three improved their performance when trained on combined data.

In summary, there is still no consensus in the literature on the optimal metrics for evaluating synthesised EEG signals. While some studies support the use of FID [53] [57] [18], others argue against it [59]. To address this gap, this thesis will utilize evaluation metrics that comprehensively assess the similarity of synthetic signals across three key domains: Fidelity, Diversity and Utility. Furthermore, while advanced models such as LSTMs [57], transformers [24], and temporal convolutional networks [50] are effective at capturing temporal characteristics, they come with significant computational costs. The variation in time window lengths across studies—from 0.5 seconds [37] to 2500 seconds [59]—also significantly impacts the quality of synthetic data, with shorter windows generally being easier to generate. To the best of our knowledge, this thesis is the first study to focus on synthesizing ioECoG signals. Although related studies have synthesized intra-cranial EEG signals [24] [55], none have specifically targeted ioECoG signals. In response to these challenges and gaps, this thesis proposes a Deep Convolutional Wasserstein Generative Adversarial Network with Gradient Penalty (DCwGAN-GP) architecture designed to generate 20-second ioECoG windows, balancing temporal complexity with computational efficiency.

4

Methods

4.1 Study Design

This section provides an overview of the methodology workflow, supported by Figure 4.1. The pipeline begins with handling the Real Data (RD) from the dataset, which contains ioECoG signals collected from patients with severe epilepsy. These signals then undergo analysis and processing at the first checkpoint: the "Quality Check." After passing this stage, the selected signals proceed to "Clustering Analysis," where they are categorized based on their data patterns. Once divided into clusters, the clustered signals move to an individualized "Generative Model", i.e., one individual generative model per data cluster, where SD is generated according to the same cluster structure defined in the "Clustering Analysis."

In the final stage, the signals are evaluated across three dimensions: Fidelity, Diversity, and Utility. The "Classification Model" assesses the latter by receiving as input the RD and SD and classifying it based on the previously established cluster labels.

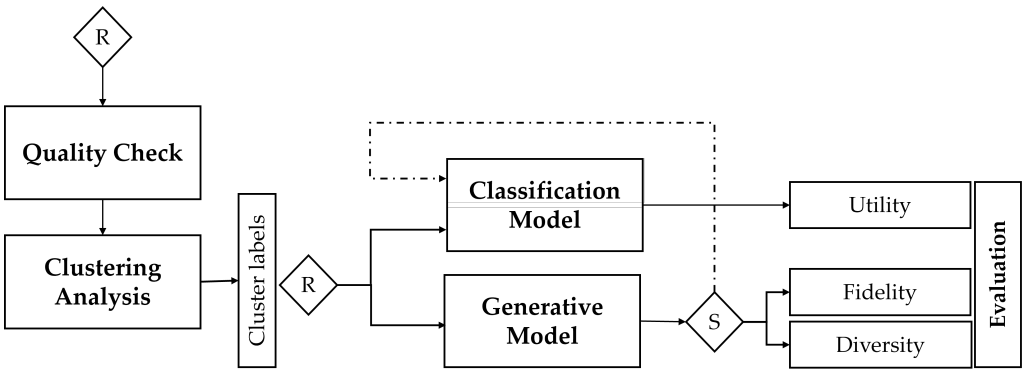

Figure 4.1: Study Design: Workflow of the pipeline carried throughout the study. "R" indicates the real signals, while "S" indicates the synthetic signals.

4.2 Sample

4.2.1 Dataset

This study encompassed data from 91 patients (mean age 14 years) who underwent epilepsy surgery from 2008 onward at the University Medical Center Utrecht (UMCU), Netherlands. This data is sourced from the Registry for Epilepsy Surgery Patients database (RESPect) and Sticthing Epilepsie Instellingen Nederland (SEIN), saved in a Brain Imaging Data Structure (BIDS) format. All patients gave informed consent for data usage. To select patients from the database, several inclusion and exclusion criteria were defined. To be included, patients had to have ioECoG recordings sampled at 2048 Hz, with photographs of the electrode placement and resected area available. Additionally, at least 1-year post-surgical seizure outcomes were required, and patients needed to be seizure-free at the latest follow-up, confirming that the non-resected tissue was free of epileptic activity. Patients were excluded if they had undergone prior brain surgery, presented with multiple epileptic foci, dual pathology, or had surgeries involving amygdala-hippocampectomy, frontal disconnection, or hemispherectomy.

4.2.2 Intra-Operative Electrocorticography Signal

During surgery, the ioECoG recordings were sampled at 2048 Hz with an anti-aliasing filter of 538 Hz, using electrode grids or strips with 4.2 mm2 platinum contact surface electrodes fixed in silicone and with an interelectrode distance of 1 cm. Importantly, propofol anaesthesia was stopped during the recording to avoid epileptiform activity suppression. Tailored surgery was guided by ioECoG to identify interictal spikes and ictiform patterns pre- and post-resection, allowing surgeons to adjust the resection area based on electrophysiological findings. The number of recording locations and the duration of recordings at each location varied for each patient, covering all areas suspected of containing epileptic tissue.

4.3 Signal Processing

In the pre-resection stage of surgery (see Section 2.1.2.1) there is a higher presence of epileptic tissue, resulting in signals with a higher content of epileptiform activity. For this reason, the signals selected for this study were from this stage.

To maintain the statistical integrity of the ioECoG signals, ensuring an authentic data generation, the signals were only trimmed and filtered, as recommended in the study of A. Delorme et al. [61]. The signals were trimmed at their midpoint, reducing the average length from 360 seconds to 60 seconds. This adjustment was made to balance computational efficiency while preserving the minimum duration necessary for the manifestation of the epileptic event (one to four minutes). As per signal filtering, a 4^{th} order Butterworth filter was applied [24, 63], with a low-cut frequency of 0.16 Hz to remove drift artifacts,

as suggested by the guidelines of the data source and similarly done in the research of K. J. Miller et al. [64], and with a high-cut frequency of 512 Hz to preserve the fast ripples present in HFO [10, 62]. Finally, an Infinite Impulse Response Notch filter at multiples of 50 Hz was also used to eliminate the powerline interference [65].

4.4 Signal Analysis and Selection

To ensure the quality of the real signals used for data generation, a thorough analysis was conducted on both the time and frequency domains of the signals.

The PSD is a fundamental tool for analyzing the frequency content of EEG signals [66], describing how its power is spread across different frequencies [67]. Given that the patients in the studied dataset were under propofol-induced anesthesia during surgery, certain frequency bands, such as the delta and alpha bands, were expected to exhibit higher power relative to others [68]. Consequently, the PSD analysis, along with the evaluation of the relative power in each frequency band, was essential to validate the signal quality [69].

Firstly, the PSD of the ioECoG signals was computed using Fast Fourier Transfrom (FFT) from the *Scipy* library [70], which transforms the time-domain signals into the frequency domain. In equation 4.1 [71], N denotes the length of both the $y[k]$ and the original time signal $x[n]$:

$$y[k] = \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{kn}{N}} \tag{4.1}$$

$$\text{Power Spectrum Density} = \frac{1}{N \times \text{sample frequency}} \times \left|y[k]\right|^2 \tag{4.2}$$

After computing the PSD with equation 4.2 [72], the relative power within each frequency band and the dominant frequency of the signal were determined. The first was calculated with the equation 4.3 [68], where the frequency band power was computed using the trapezoidal rule, which estimates the area under the power spectrum curve within each frequency band. The dominant frequency of each signal was simply determined as the frequency corresponding to the maximum value in the PSD.

$$\text{Relative Band Power} = \frac{\text{Frequency Band Power}}{\text{Total Frequency Power}} \tag{4.3}$$

It has been stated that EEG signals recorded during sleep should exhibit a relative power in the delta band greater than 0.18 Hz and a dominant frequency at or below 5 Hz [29]. Based on these criteria, the signals were systematically evaluated and outlier signals were identified by comparison with these benchmarks.

4.5 Clustering

Traditionally, clinicians during epilepsy surgery visually inspect the ioECoG signals with a low-pass filter up to 70 Hz to differentiate between epileptiform and non-epileptiform activity [10], thus distinguishing the injured and healthy tissue. However, ML studies present in the literature consider frequencies above the 70 Hz range for this distinction, the HFO [62], [32]. Additionally, during the visual inspection of the signals during Section 4.4, there were found several distinct data patterns that were more complex than the clinicians's binary classification. These findings raised the doubt that the labels provided by the clinicians might not fully capture the complexity of the data. For this reason, a naive, unsupervised clustering analysis was conducted. This data-driven approach allowed to objectively explore the underlying patterns in the data, thereby identifying distinct clusters of signals based solely on their inherent characteristics. It not only provided this better understanding of the data, essential for labelling the signals, but also facilitated the development of an individualized generative process. By dividing the signals into distinct types and providing this information as individual labels to their specific generative models, we aimed to enhance the model's ability to generate more accurate and representative synthetic signals.

4.5.1 Feature Extraction

For the clustering analysis 33 distinct features were extracted from the signals using the *TSFEL* [73], Numpy [74], and *Scipy* [70] libraries. These features were selected to capture the various characteristics of the signals across four domains: temporal, frequency, time-frequency, and non-linear. The temporal features were primarily derived from the methodologies of the study conducted by Rasheed et al. [63] and Du et al. [75], given their performance in clustering epileptic EEG signals, while the time-frequency features were based on the work of Chaibi et al. [15] for its purpose in distinguishing epileptiform activity, both similar to the tasks conducted in this thesis. The frequency and non-linear features were specifically computed for this study, based on the signal analysis detailed in Section 4.4. Key features are explained in more detail in Table 4.1.

4.5.1.1 Feature Processing and Selection

Following the extraction of features, they underwent two preprocessing steps. First, they were standardized using the *StandardScaler* from *Scikit-Learn* [80], to ensure that differences in scale did not disproportionately influence their relative importance [81]. Scaling the features with the *StandardScaler* means removing its mean and scaling it to unit variance [81]. This approach follows the work of Guerrero et al. [82] and D.K et al. [83], where the latter identified standard scaling as one of the most effective scaling techniques for detecting epileptic events in EEG signals. Next, features with a correlation above 95% were removed to reduce the complexity and redundancy of the dataset [84].

Table 4.1: Feature Extraction. Description of example features extracted for the Clustering Analysis.

Domain	Feature	Description
Temporal	Zero Crossing Rate	Rate of sign changes in the signal [76].
	Shannon Entropy	Measure of uncertainty of a random variable [77].
Non-Linear	Hurst Exponent (H)	Quantification between $[0, 1]$ of the long term self-similarity of the signal. $H > 0.5$ are considered as persistent series [78].
	Autocorrelation	Measurement between $[-1, 1]$ of the correlation between a signal and its delayed versions . Values near one indicate higher autocorrelation [79].
Frequency	Spectral Spread	Measurement of how the spectrum is distributed around its mean value [73].
	Delta Relative Power	Delta band power relative to total power (equation 4.3) [68].
	Dominant Frequency	Frequency with the highest power in the PSD.
	Mean FFT	Mean amplitude of the FFT of the filtered signal, using a 30-70 Hz bandpass for spike detection [15].
	Mean Peak Amplitude in the 1-2Hz Range	Average peak amplitude within the 1-2 Hz range in the FFT, motivated by the findings in section 4.4.
Time-Frequency	Normalized Number of Peaks	Number of peaks in the time-frequency map, normalized by the number of elements [15].

Finally, to identify the most discriminative features, the signals were grouped into two categories given by the clinicians' labels: healthy and injured tissue. Then, the mean value of each feature was computed per group, for the two groups and 33 features. After their computations the mean values were compared between the groups.

4.5.2 Cluster Quantification

Given the unsupervised approach taken in the clustering analysis, two well-known methods were employed with the *Scikit-Learn* library [80] to determine the optimal number of clusters to use in the clustering algorithm: the elbow method and silhouette analysis [85].

- **Elbow Method:** In this method, the K-means algorithm is applied for multiple values of k (the number of clusters), calculating the dispersion of the sum of squared distances between data points, the Within-Cluster Sum of Squares (WCSS), and their respective cluster centroids. The optimal number of clusters is identified by plotting the WCSS against the number of clusters. The point where the rate of decrease in dispersion slows marks the optimal value for k, known as the elbow point [85].

- **Silhouette Analysis**: In this analysis the silhouette coefficient is computed to evaluate the quality of clustering by considering two characteristics of each sample: the intra-cluster distance (a), average distance between a sample and the other points

in the same cluster, and the nearest-cluster distance (b), average distance between a sample and all points in the closest cluster that the sample is not part of. The Silhouette Coefficient (s) for a single sample is then calculated as 4.4 [86]:

$$s = \frac{b - a}{\max(a, b)} \tag{4.4}$$

In which a score near 1 means that the sample is assigned to the correct cluster [87], near 0 may indicate an overlapping cluster or a wrong assignment [87] [80], and a value near -1 means that the sample is in the wrong cluster [87]

The optimal number of clusters was then determined by analyzing the elbow point from the Elbow Method and identifying the number of clusters that corresponded to the highest average silhouette score from the Silhouette Analysis.

4.5.3 Clustering Analysis

After selecting the most discriminative features and determining the optimal number of clusters, the Agglomerative Clustering algorithm from *Scikit-Learn* [80] was employed to divide the signals into distinct patterns. This is an unsupervised algorithm that utilizes a "bottom-up" approach where each data point starts as an individual cluster. Clusters are then recursively merged based on a chosen linkage criterion, such as single linkage (minimum distance), complete linkage (maximum distance), average linkage, or Ward's method (minimizing variance). This process continues until all points are combined into a single cluster or a specified number of clusters [88]. In this study the Ward's method was used.

To further examine the different distribution of the signals per cluster and assess the relevance of the selected features, two 3D scatter plots of the clustered data and the clinicians' labels were generated. The axes of the 3D plot represent the different features utilized in the clustering process, allowing for a visual assessment of how the features contribute to the formation of the clusters. Finally, the clusters were labeled having in consideration the nature of the signals and by comparing the clustering results with the clinicians' classifications through the percentage of alignment between labels.

4.6 Generative Model

This section introduces the structure of the proposed generative model: DCwGAN-GP. It first outlines the input data used for the model, followed by a description of its architecture, the parameters used for its development and the training process. This generative framework was conducted with the use of *Pytorch* [89], utilizing a Graphics Processing Unit (GPU) NVIDIA RTX 6000 Ada Generation [90].

In order to generate synthetic data specialized in each data cluster, an individual generative model was created per each cluster. The architecture and training of the

individual models was conducted identically, except for the batch size utilized in the last cluster. Therefore, during this section, the procedure of the creation of the Generative Model will be applied to all models.

4.6.1 Data Preparation

After the initial processing of the ioECoG data, the signals, now divided into clusters, required further preparation before being fed into the generative model. This preparation involved segmenting and downsampling the signals, primarily for computational efficiency. According to the literature review in Section 3, the most common time window for generative signals is 4 seconds, with a typical sampling frequency of around 256 Hz. In this study, the 60-second signals were segmented into 20-second windows and downsampled to 512 Hz, balancing the computational costs with the generation of realistic samples. Finally, the signals were normalized by dividing them by its maximum value and then transformed into tensors to serve as input for the generative model.

4.6.2 Model Architecture

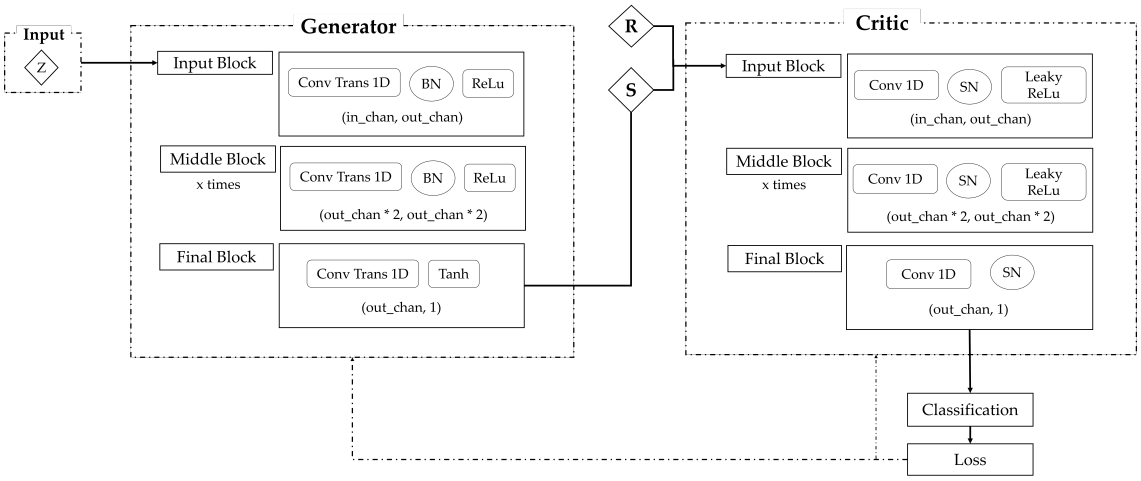

Figure 4.2: Generative Model's Architecture. Each Neural Network is comprised of three types of blocks: input, middle and final block. The input and final block are constituted by one layer, while the middle block was constituted by several layers. The number of layers contained in the the middle block was defined through the optimization of the model.

After preparing the input data, the model architecture was designed, as shown in Figure 4.2. Both the Generator and Critic were constructed using 1D Convolutional Neural Networks (CNNs) and trained using the Wasserstein Distance loss function with a Gradient Penalty. Although CNNs are typically used for image processing, where the data is two-dimensional, 1D Convolutional Layers are proven to capture complex patterns within the EEG signals, which are one-dimensional [21].

The Generator starts by receiving as input a latent vector of size 100, Z, which is then processed through a series of 1D Transposed Convolutional Layers, configured with a kernel size, stride, and padding of (4, 2, 1). These layers upsample the input data, enabling the Generator to produce more realistic samples [21]. Batch normalization (BN) and ReLU as the activation function are utilized between layers. In the final layer of the Generator, a hyperbolic tangent (Tanh) activation function is applied, generating the synthetic samples: G(z). In turn, the Critic receives both the real and fake samples as input, and are processed through 1D Convolutional Layers, also configured with a kernel size, stride, and padding of (4, 2, 1). Spectral Normalization is applied to ensure stability during training [91], while Leaky ReLU with alpha as 0.2 [22] serves as the activation function. Finally, the critic outputs a classification score that differentiates between real and fake samples, backpropragating its loss into the networks.

4.6.3 Model Training and Tuning

The training process for the DCwGAN-GP models was organized into several key steps, each designed to iteratively improve the Critic and Generator. The steps are detailed as follows:

1. **Start of the Training Loop**

   The training loop iterates over a predefined number of epochs. For each epoch, the loop processes batches of real data with a specified size.

2. **Critic Training Loop**

   Since Critics train faster than the Generators [22], the Critic in this model was trained five times for each update of the Generator, similarly to the work of Carrle et al. [23], ensuring that it effectively distinguishes between real and fake data.

   a) At the beginning of each Critic update, the Critic's gradients are initialized to zero to prevent accumulation from previous iterations [92].

   b) A latent tensor, z , is sampled from a standard normal distribution $z \sim \mathcal{N}(0, I)$, with a shape (batch size x latent dimension).

   c) Then, the Generator uses this tensor to generate fake data ($G(z)$).

   d) The Critic processes the real, (x), and fake data, ($G(z)$), outputting a score for each: $C(x)$ and $C(G(z))$.

   e) The Gradient Penalty, 4.5, is computed with the third term of the equation 2.3:

$$\text{Gradient Penalty} = \lambda \mathbb{E}_{x \sim p_{\text{data}}(x)} \left[(\|\nabla C(\hat{x})\|_2 - 1)^2 \right] \tag{4.5}$$

   f) The Critic's loss is then calculated as equation 2.3:

$$L = \mathbb{E}_{x \sim p_g} \left[C(G(z)) \right] - \mathbb{E}_{x \sim p_{\text{data}}(x)} \left[C(x) \right] + \text{Gradient Penalty} \tag{4.6}$$

This loss encourages the Critic to maximize the difference between its scores on real and fake data while penalizing large gradients.

g) The Critic's loss is backpropagated, and its parameters are updated using the Adam optimizer.

h) Repeat 5 times.

3. **Generator Training**

a) Similarly to the Critic, the generator's gradients are initialized to zero.

b) The Generator then produces a new batch of fake data, $G(z)$, using the previously created latent tensor, z.

c) The Critic evaluates this new fake data, outputting a score: $C(G(z))$.

d) The Generator's loss is calculated as 4.7:

$$\text{loss}_G = -\mathbb{E}[C(\text{G(z)})] \tag{4.7}$$

This loss pushes the Generator to improve the quality of the fake data, so that the Critic scores it closer to the real data.

e) The loss is backpropagated, and the Generator's parameters are updated using the Adam optimizer.

f) The losses for both the Critic and Generator are stored for later analysis.

4. **End of the Training Loop**

At the end of each epoch, the learning rate schedulers for both the Generator and Critic are updated to adapt the learning rates during training.

After training the networks, the model underwent several rounds of tuning to achieve optimal performance. The final optimized versions of all models involved the tuning and training of fifty-one models. This last version featured six blocks in the Generator and five in the Critic for all individual models. The final training parameters are summarized in Table 4.2.

Table 4.2: Training Parameters for DCwGAN-GP. The training parameters were consistent across all individual models, with the exception of the batch size, which was reduced to 30 in the final model to optimize performance, and the number of samples, which varied accordingly.

Training Parameters	
Number of Epochs	2000
Batch Size	40*
Learning Rate	0.00005
Critic Interactions	5
Gradient Penalty Lambda	10
Latent Dimension	100
Number of samples	Number of data points in each cluster

4.7 Evaluation Metrics

This section details the evaluation metrics used to assess the quality of the generated ioECoG signals. These metrics are categorized into the domains of Diversity, Fidelity and Utility, as discussed in Section 2.2.2. Before evaluating the similarity between real and synthetic signals, the real signals were downsampled to 512 Hz and normalized to match the structure of the generated signals.

4.7.1 Fidelity Analysis

Fidelity is a metric essential for assessing the similarity between synthetic and real ioECoG signals, ensuring that the synthetic data accurately and reliably reflects the characteristics of the real data, while guarantying that the synthetic data is not merely a replica. To achieve this, various metrics were applied in the time, frequency, and time-frequency domain to assess these qualities.

4.7.1.1 Time Domain

To evaluate the synthetic signals in the time domain, the mean and standard deviation (STD) of both real and synthetic signals were calculated and visualized. Additionally, the WD was computed, quantifying the differences between the probability distributions of the real and synthetic signals. It calculates Earth Mover's Distance between two distributions of real and synthetic data, representing the cost required to transform one probability distribution into another [93]. For its calculation the signal data was divided into bins, and the probability distribution was computed. Then the Wasserstein Distance can be calculated as equation 4.8 [93]:

$$D_{WD}(P\|Q) = \inf_{\pi \in \Gamma(P,Q)} \int_{\mathbb{R}\times\mathbb{R}} |x - y| d\pi(x, y) \tag{4.8}$$

where $\Gamma(P, Q)$ is the set of all joint distributions whose marginals are P and Q, respectively. For any given x, $P(x)$ describes the probability of P at the point x, and similarly, $Q(x)$ does the same for Q [93].

4.7.1.2 Frequency Domain

To evaluate the frequency content of the synthetic signals, their PSD were calculated following the methodology described in Section 4.4. This analysis addresses the Fidelity domain by ensuring that the synthetic signals maintain similar spectral characteristics to the real signals. After calculating the PSD, a plot was generated to visually compare the power distribution across frequencies between the synthetic and real signals.

4.7.1.3 Time-Frequency Domain

To assess the dynamic changes in frequency content over time, scalograms of the real and synthetic signals were computed using the Morlet Wavelet Transform. A scalogram

provides a time-frequency representation by plotting the absolute value of the Continuous Wavelet Transform (CWT) over time and frequency [94]. With the Wavelet Transform, the signals are decomposed into shifted and scaled versions of a wavelet, representing the data across several scales [95]. The choice of the Morlet Transform was based on its ability to detect and localize fine details, providing a good time localization [96].

Several metrics were used to quantify the similarity between real, S_{real}, and synthetic, $S_{\text{synthetic}}$, scalograms:

- **Pearson Correlation Coefficient**: Quantifies the linear relationship between two variables [97]. In this study it was calculated with the *Scipy* library [70]. It can be calculated as 4.9 [97].

$$Coefficient = \frac{\sum_{i=1}^{n}(S_{\text{real},i} - \overline{S}_{\text{real}})(S_{\text{synthetic},i} - \overline{S}_{\text{synthetic}})}{\sqrt{\sum_{i=1}^{n}(S_{\text{real},i} - \overline{S}_{\text{real}})^2}\sqrt{\sum_{i=1}^{n}(S_{\text{synthetic},i} - \overline{S}_{\text{synthetic}})^2}} \tag{4.9}$$

- **Cosine Similarity**: Measurement of the cosine of the angle between two vectors, providing an indication of their similarity in orientation [98]. It is calculated as 4.10 [99].

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{N} S_{\text{real},i} \times S_{\text{synthetic},i}}{\sqrt{\sum_{i=1}^{N} S_{\text{real},i}^2}\sqrt{\sum_{i=1}^{N} S_{\text{synthetic},i}^2}} \tag{4.10}$$

- **Structural Similarity Index (SSIM)**: Metric that considers changes in structural information, s, luminance, l, and contrast, c, between a reference image, x, and a corrupt version of the same image, y, [100]. It can be calculated with the equation 4.12 [100]. This metric was computed with the *Scikit-Learn* Image Processing library [101].

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{4.11}$$

$$\text{SSIM}(x,y) = l(x,y) \cdot c(x,y) \cdot s(x,y) \tag{4.12}$$

- **Mean Squared Error (MSE)**: Metric to evaluate the error between a reference image, the real scalogram, and its corrupt version, the synthetic scalogram. Values closer to zero indicate a lower error. It can be calculated as equation 4.13 [102].

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(S_{\text{real},i} - S_{\text{synthetic},i}\right)^2 \tag{4.13}$$

4.7.2 Diversity Analysis

To evaluate how the synthetic data diverges and its resemblance to the diversity of the real data, two dimensional reduction techniques to visualize high-dimensional data [103] were utilized: PCA and the t-SNE [24]. The first measures how much of the variance in

the data can be explained by each principal component [104], while the latter preserves the relationships between data points [103].

4.7.3 Utility Analysis

Finally, to evaluate the utility of the synthetic data, a classification model was developed. The classifier was trained on three different datasets: real data, synthetic data, and a combination of both. The goal was to determine whether the synthetic data could enhance the classifier's ability to correctly identify and differentiate between signal classes. Improved performance with the augmented dataset would indicate that the synthetic data is not only faithful and diverse but also useful in practical applications.

4.7.3.1 Train and Test Split

The features used in the classification process were identical to those utilized in the clustering algorithm in Section 4.5.1. The datasets were then split into training and testing sets, with 70% allocated for training and 30% for testing. Stratified sampling was applied to ensure class proportions were maintained in both the training and testing sets.

4.7.3.2 Model Training

Several ML classifiers were trained on the training datasets to develop predictive models. All models were created using the *Scikit-Learn* library [80]. The classifiers included:

- **Random Forest:** An ensemble algorithm that combines multiple decision trees to make a prediction. It builds multiple trees from random data subsets, then combines their predictions for improved accuracy in classification and regression [105].
- **Logistic Regression:** Similar to the Linear Regression, it calculates a weighted sum of the input features and applies the logistic function to the result, producing an output between zero and one [106].
- **Support Vector Machines (SVM):** A linear classifier that optimally divides the data into two separate groups, creating a hyperplane in a higher-dimensional space [107].
- **Neural Network Multi-Layer Perceptron:** A type of artificial NN comprised of multiple layers of neurons that can capture the non-linear relationships in the data [108].
- **Gradient Boosting:** An ensemble method that combines weak learners sequentially, with each model correcting the errors of the previous one. Particularly effective for tabular supervised tasks, including both classification and regression [109].
- **Balanced Random Forest:** An adaptation of Random Forest, specifically designed to handle imbalanced datasets effectively [110].

These models were selected to address the imbalanced, tabular, and non-linear nature of the dataset. RF, Gradient Boosting, and NNs were chosen for their ability to capture complex non-linear patterns, with Gradient Boosting being particularly effective in handling tabular data. The Balanced RF was included to manage class imbalance, while Logistic Regression and SVM served as simpler baselines for comparison.

4.7.3.3 Model Evaluation

To ensure the models were trained effectively, the Stratified K-Fold Cross-Validation from the *Scikit-Learn* [80] library was employed, with k=10, following the methodology of [50]. This method involves splitting the training data into 10 subsets, training the model on 9, and validating it on the remaining one. This cycle is repeated 10 times for each fold and with each subset serving as the validation set once. In the end, the average performance across all folds is computed [106]. To minimize bias from any inherent order (i.e., class labels), the data was randomly shuffled before being split into folds, ensuring that each fold contained a representative sample of the dataset. This was achieved by setting the shuffle parameter to "True".

After training, the models were evaluated on both the training and testing datasets. This evaluation, facilitated by the train-test split, is essential to ensure that the model generalizes well on unseen data, i.e., the test data, and does not overfit to the training data [106]. To assess the models' performance, metrics such as Accuracy, F1-Score, Precision, and Recall were calculated and compared between classifiers.

5

Results

This section presents the results obtained throughout the study to accomplish the objectives outlined in Section 1.2. First, the outcomes of the Clustering Analysis (Section 4.5.3) are presented, highlighting the patterns identified in the real ioECoG data. Next, the evaluation of the generated data is described, focusing on its similarity to the real data and its utility in augmenting the dataset for improved classification.

5.1 Cluster Analysis

The analysis of the clusters revealed the underling patterns present in the ioECoG data. These patterns were identified through the Agglomerative Clustering algorithm and several discriminative features. Although 33 features were computed to distinguish the different data types, eight showed superior performance in cluster aggregation: hurst exponent, autocorrelation, mean peak amplitude in the range 1-2 Hz, spectral spread, delta relative power, dominant frequency, number of peaks depicted in the time-frequency representation, and the mean of the absolute value of the FFT spectrum of the segment filtered between 30 to 70 Hz (see Table 4.1).

This section will describe the type of patterns identified through the clustering analysis, along with their defining traits.

5.1.1 Cluster Quantification

Figure 5.1 compares the Elbow and Silhouette methods to determine the optimal number of clusters. In the Elbow Method, the WCSS decreases sharply until four clusters, where the decline slows, forming the elbow point. In the Silhouette Analysis, the highest silhouette score (0.49) occurs at two clusters, with the score steadily decreasing to around 0.20 after three clusters. Based on these findings, the Elbow Method and silhouette score analysis led to the selection of three clusters. It is important to note that, although the highest silhouette score was achieved with two clusters, the algorithm did not distinguish between physiological (healthy/injured) tissue at this level. However, with k=3, it successfully made this distinction.

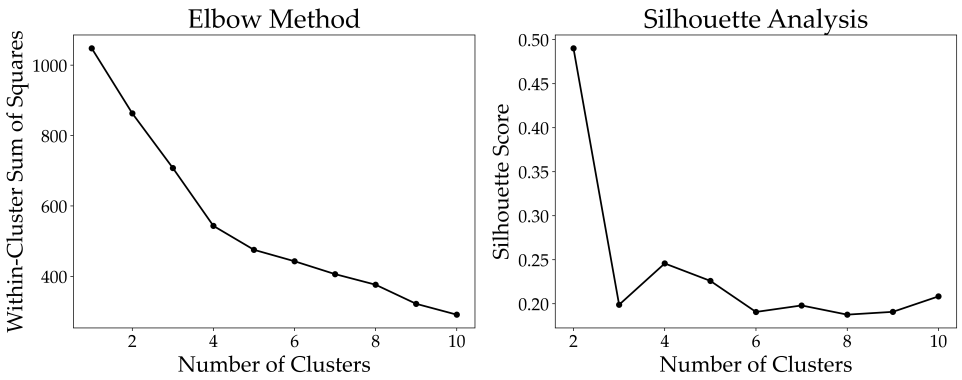

Figure 5.1: Elbow and Silhouette Analysis. The first identifies the optimal number of clusters by locating the inflection point in the Within-Cluster Sum of Squares graph, while the second selects the number of clusters that yields the highest silhouette score.

5.1.2 Cluster Assignment

After categorizing the real signals into three clusters, a 3D scatter plot was generated to visualize the distribution of data points, alongside another plot based on the clinician's labels for comparison. The left side of Figure 5.2 shows the data points separated into three clusters: Cluster 0 (orange), Cluster 1 (purple), and Cluster 2 (blue). The right side displays the clinical binary labels, distinguishing Healthy (blue) and Injured (purple) data points. Both plots share the same axes: mean peak amplitude in FFT (1-2 Hz) on the x-axis, autocorrelation on the y-axis, and delta relative power on the PSD on the z-axis.

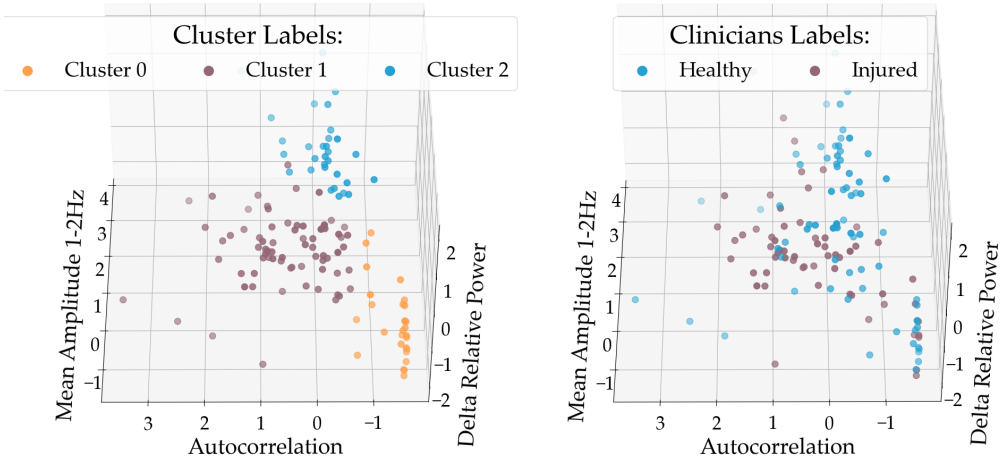

Figure 5.2: 3D Scatter Plot Comparison. The left side of the Figure displays the data points labeled by the clustering algorithm, while the right side shows the binary labels assigned by the clinicians.

In the clustering results, Cluster 0 is concentrated in the negative regions of all axes, while Cluster 1 occupies the center with lower x-axis values, and Cluster 2 is positioned at higher z-axis values, scattered along positive x-axis values. Cluster 1 has the highest

density, while Clusters 0 and 2 have similar densities. Despite overall separation, some overlap is observed, particularly between Clusters 0 and 1, and between Clusters 1 and 2, with a few misclassified points near the boundaries. Regarding the clinician-labeled plot, Healthy data points are more concentrated in the positive regions of the x- and z-axes, similar to Cluster 2, while Injured points are spread more towards negative y-axis values, overlapping with Healthy points. A closer comparison of the clusters with the clinician labels reveals that Cluster 2 aligns mostly with the Healthy cluster and Cluster 1 with the Injured, while Cluster 0 comprises a combination of both types. Specifically, Cluster 0 contained 57.14% Healthy and 42.86% Injured data points, Cluster 1 had 38.89% Healthy and 61.11% Injured, and Cluster 2 contained 83.87% Healthy and 16.13% Injured.

After clustering the data points, their signals and PSDs were analyzed. Figure 5.3 shows a representative signal from each cluster, followed by their respective PSDs and the mean PSD with the STD.

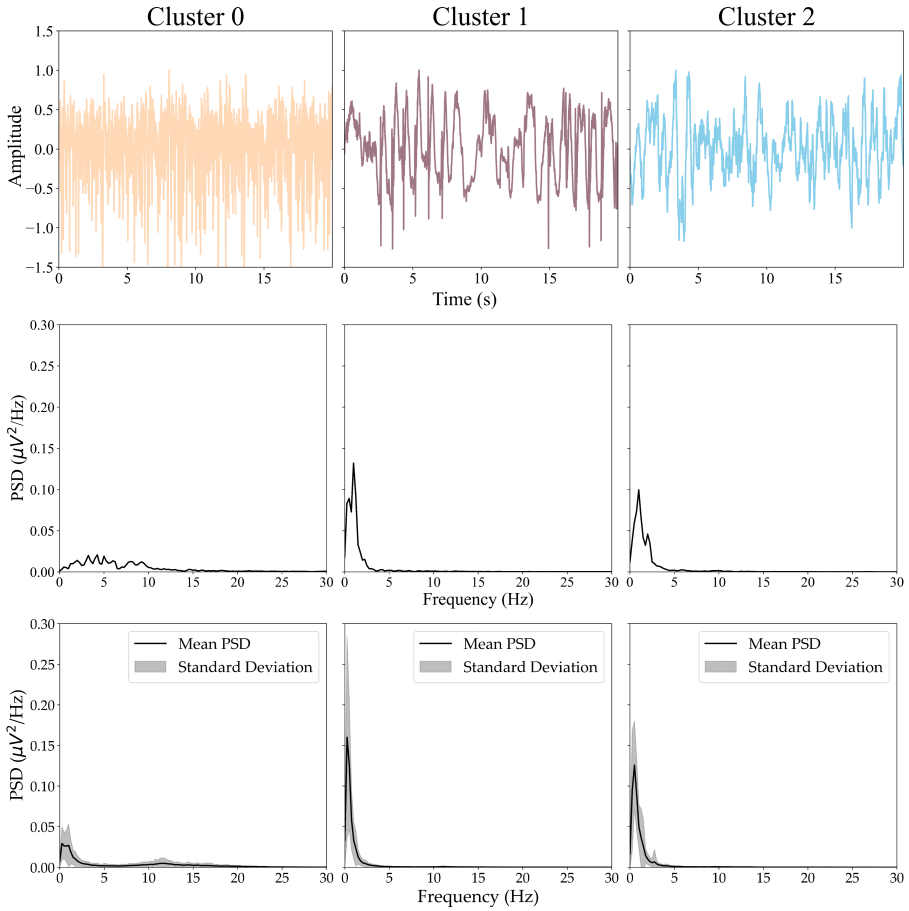

Figure 5.3: Characteristics of Representative Signals. This Figure showcases three representative signals of each cluster, followed by their Power Spectrum Density and by the mean and standard deviation of the of the Power Spectrum Density of each cluster. The signals were normalized, therefore the amplitude is in arbitrary units.

The signal from Cluster 0 is characterized by irregular, noisy fluctuations, reflected in a low PSD with no dominant frequency peaks. However, the mean PSD shows a

small peak in the delta EEG band and a subtle fluctuation in the alpha band, with a low STD indicating minimal variability across the signals. In contrast, Cluster 1 displays a signal with structured patterns, a smoother profile and less fluctuations, characterized by its strong peak in the delta band in its PSD. In addition, the larger STD in the mean PSD suggests greater signal variability within the cluster. Finally, Cluster 2 signals' are relatively noisy, but show more organized fluctuations compared to Cluster 0 and less than Cluster 1. The PSD also reveals a strong peak in the delta band, similar to Cluster 1, with a lower STD, suggesting more consistent signal patterns within the cluster. Following the clustering analysis, the comparison with the clinician's labels and the analysis of the time and frequency domains of the signals assigned to each cluster, the clusters were categorized into: Cluster 0 – Noisy, Cluster 1 – Injured/Epileptogenic, and Cluster 2 – Healthy/Non-Epileptogenic.

5.2 Evaluation of Synthetic Data

The real signals serve as the foundation for the generative process, where the aim of the generative model is to replicate the physiological characteristics present in the original dataset. Accordingly, this section will conduct a comparative evaluation of the real and generated signals across the three clusters within the previously defined evaluation domains.

5.2.1 Fidelity Analysis

5.2.1.1 Time Analysis

Figure 5.4 presents the comparison between the mean of the time domains for both the real and synthetic Cluster 0, Cluster 1, and Cluster 2. The waveform of the signals' mean is kept throughout the three synthetic clusters, with Cluster 0 having a slightly higher amplitude than its real version and being visibly noisier.

Figure 5.5 compares the WD between three data pairings: Real-Real (RR), Real-Synthetic (RS), and Synthetic-Synthetic (SS). The WD measures the similarity between distributions, that is, the cost required to transform one probability distribution into another. Thus, lower values indicate higher similarity between the two distributions. Overall, the differences in the similarity across the different combinations of the data are relatively small, with WD values ranging from 0.022 ± 0.017 (C0), 0.034 ± 0.023 (C2) and 0.055 ± 0.036 (C1) in the SS pairing, up to 0.038 ± 0.028 (C0), 0.051 ± 0.040 (C2) and 0.072 ± 0.053 (C1) in the RR. Additionally, a clear tendency is observed, with the RR pairing exhibiting the highest WD in the three clusters, followed by RS, and the lowest values appearing in the SS combination. Throughout this tendency, the clusters within each pairing also follow their own distinct trend in WD variability. Cluster 1 consistently displays the highest WD values across all pairings, followed by Cluster 2, with Cluster 0 always showing the lowest values.

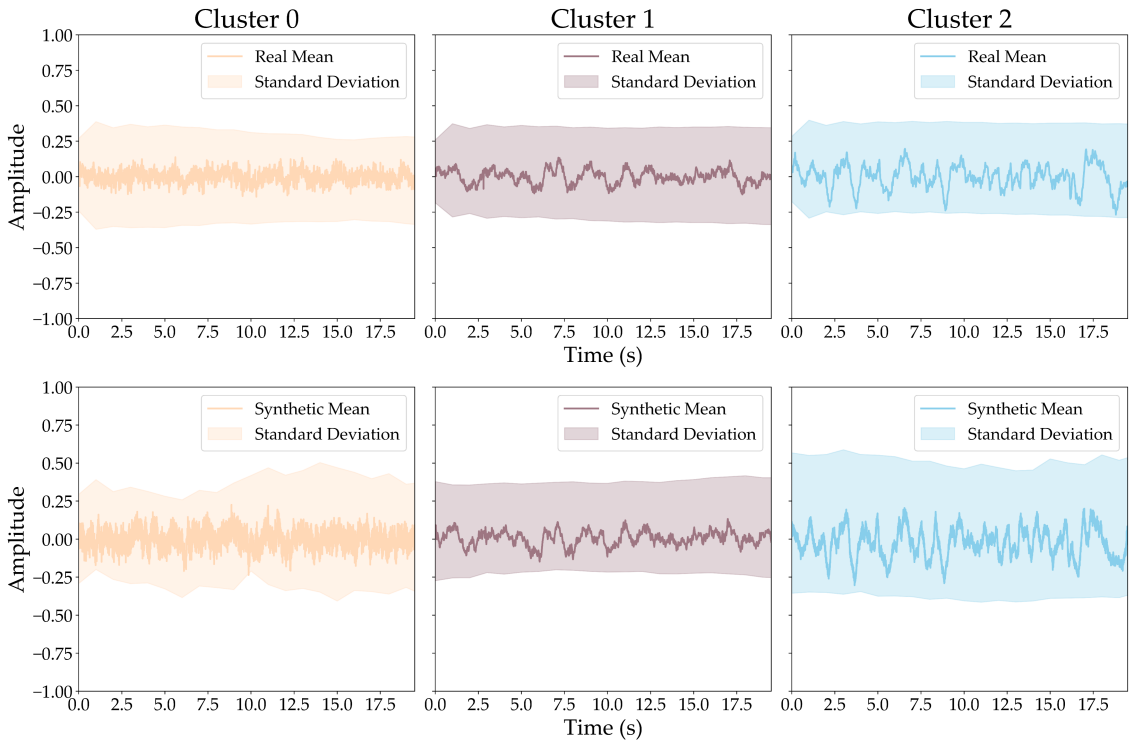

Figure 5.4: Time Domain Comparison. The Figure illustrated the comparison between the mean and standard deviation of the real and synthetic signals. The first line is referred to the real signals, while the second is referred to the synthetic. The signals are normalized, therefore the amplitude is in arbitrary units.

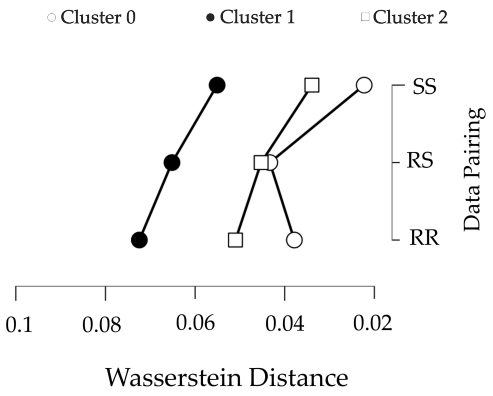

Figure 5.5: Comparison of the Wasserstein Distance between different data pairing. In this case, this metric is being compared between the real and synthetic (RS) signals, and within the real (RR) and synthetic (SS) signals separately. Each cluster is represented by a geometric Figure: Cluster 0 by a white circle, Cluster 1 by black circle and Cluster 2 by a white square.

5.2.1.2 Frequency Analysis

Figure 5.6 evaluates the clusters in their frequency domain. It compares the mean PSD and the respective STD of the real and synthetic signals across the three clusters, over a frequency range of 0 to 30 Hz for better visualization. Across all clusters, the synthetic signals exhibit similar low-frequency behavior to the real signals, with power concentrated in the delta band. However, the Fast Delta (2-4 Hz) contribution is consistently higher in the synthetic signals, particularly in Cluster 0 and Cluster 2 with values of 18.12% and 18.77% respectively, compared to 7.93% and 7.03% of their real versions. Additionally, the synthetic signals consistently have a lower STD across all clusters, indicating less variability compared to the real signals.

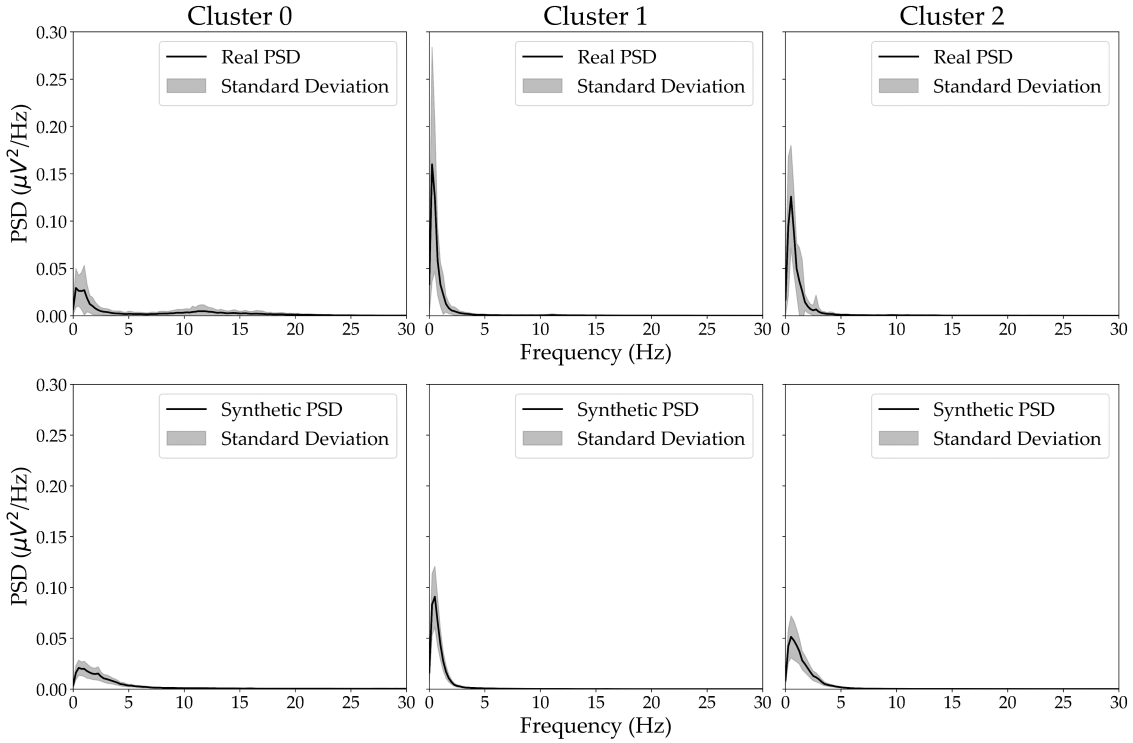

Figure 5.6: Comparison of the mean Power Spectrum Density and respective Standard Deviation between real and synthetic clusters. The first line is referred to the real signals, while the second is referred to the synthetic signals.

5.2.1.3 Time-Frequency Analysis

Figure 5.7 showcases four metrics - Cosine Similarity, Structural Similarity Index (SSIM), Pearson's Correlation and Mean Squared Error (MSE) - computed to assess the similarity of synthetic data within the time-frequency domain by comparing the scalograms of different data points. These metrics were evaluated through the same data pairing utilized in the computation of WD: RR, RS and SS.

The three clusters transition from moderate Cosine Similarity values in the RR pairing

(average value of 0.729 ± 0.037 between the three clusters) to a higher similarity in the SS combination (average value of 0.795 ± 0.005) indicating a stronger resemblance between the synthetic scalograms compared to their original counterparts. SSIM and Pearson's Correlation follow a similar trend, but with Cluster 1 and Cluster 2 showing higher values compared to the Noisy Cluster. Additionally, the Noisy Cluster shows an increase in inter-similarity in the SS combination, with a 0.180 ± 0.051 increase in SSIM and nearly doubling its Pearson's Correlation value (0.275 ± 0.119 increase) compared to the RR combination. Across the three clusters, the Mean Squared Error demonstrates an inverse trend, with an average decrease of 0.182 ± 0.052 from the RR to the SS combination. In the direct comparison of real and synthetic scalograms, Cluster 1 and Cluster 2 maintain moderate-to-high similarity in the Pearson's Correlation, SSIM and Cosine Similarity, while Cluster 0 lowers the overall average of the first two, with mean values of 0.566 ± 0.165, 0.542 ± 0.147, and 0.727 ± 0.036 for these metrics, respectively. Additionally, the three clusters exhibit moderate Mean Squared Error values in the RS pairing, with an average of 0.499 ± 0.078, where Cluster 1 shows the lowest error.

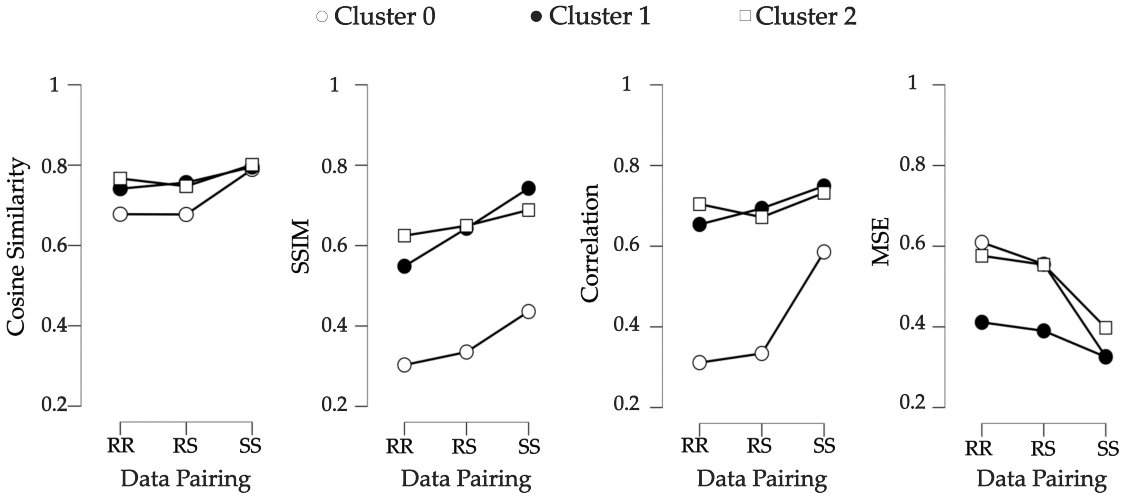

Figure 5.7: Similarity Metrics in the Time-Frequency Domain between different data pairing: real-real (RR), real-synthetic (RS) and synthetic-synthetic (SS). Each cluster is represented by a geometric Figure: Cluster 0 by a white circle, Cluster 1 by black circle and Cluster 2 by a white square. The similarity metrics are enconded as Cosine Similarity; SSIM - Structural Similarity Index Measure; Correlation - Pearson's Correlation; MSE - Mean Squared Error.

5.2.2 Diversity Analysis

Figure 5.8 evaluates the diversity of the data through the PCA and t-SNE visualizations for both real and synthetic signals throughout the three clusters. In the PCA results, the synthetic data consistently overlaps with the real data across the clusters, replicating the overall global structure of the real data. Additionally, the synthetic data is concentrated in the center region of the plot, whereas the real data is more spread out, indicating a

lack of diversity in the synthetic data. Considering the t-SNE plots, which focus on local relationships, the synthetic signals are more scattered and less grouped compared to the real signals, indicating that the synthetic signals could not capture the local relationships that are present in the real data.

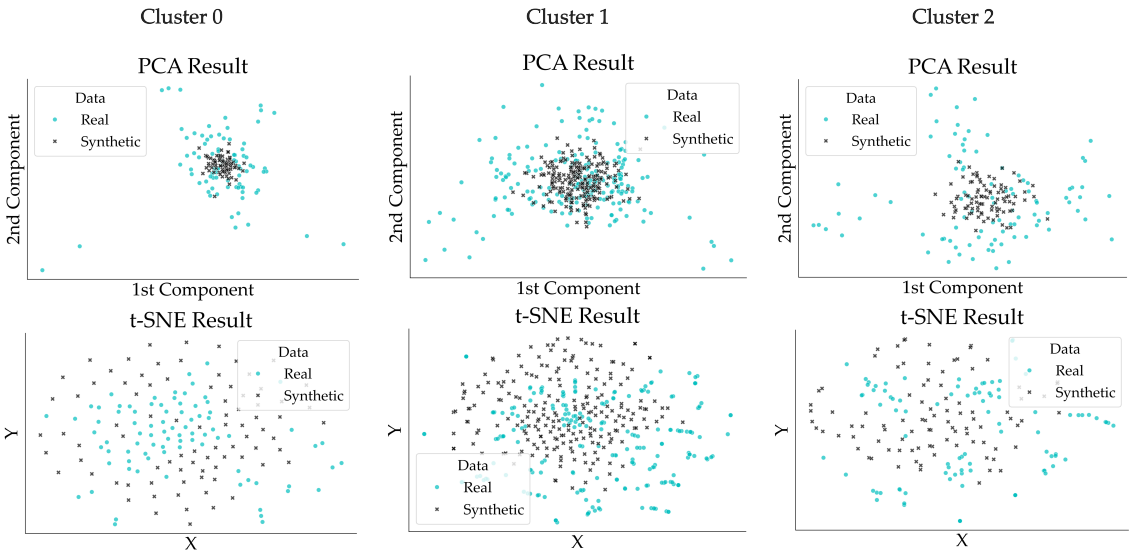

Figure 5.8: Diversity Evaluation. The first line of the figure is correspondent to the PCA evaluation, while the second line is correspondent to the t-SNE evaluation. The columns correspond to each cluster. The synthetic data is represented by black dots, while the real data is represented by blue.

5.2.3 Utility Analysis

The four Tables presented below, 5.1, 5.2, 5.3 and 5.4, showcase the performance of the six classifiers trained on real, synthetic, and combined (real and synthetic) data. The value scores are similar across the four metrics: Accuracy, Precision, Recall and F1-score. When trained on real data, RF, NN and Gradient Boosting achieved 100% scores throughout all metrics, but experienced test drops from 10% to 17%. Logistic Regression maintained consistent scores across both training and test datasets, while the Balanced Random Forest reached a 6% decrease and SVM had a 2-3% drop in performance between training and test sets. For synthetic data, all classifiers achieved 100% scores across all metrics on both the training and test sets. When trained on the RS data, the classifiers improved their test performance compared to training on real data alone, reaching from 5% to 12% increases. RF, NN, and Gradient Boosting maintained 100% training scores, having test reductions around 4% to 7%. Logistic Regression and SVM scored around 94% in the four metrics on the training set, while Balanced RF performed at around 97%. These three classifiers showed a test performance drop from 1% to 5%. Regarding mean cross-validation accuracy, all classifiers achieved 1.00 ± 0.00 on synthetic data, while the performance on real data ranged from 0.83 ± 0.05 to 0.87 ± 0.04. Training on the RS data yielded cross-validation

scores of 0.93 to 0.94 ± 0.03-0.04 for RF and Balanced RF.

Table 5.1: Classifiers' Accuracy. Presentation of Train, Test, and Mean Cross-Validation Accuracies across all classifiers.

Classifier	Train			Test			Mean Cross Validation		
	R	S	RS	R	S	RS	R	S	RS
Random Forest	1.00	1.00	1.00	0.89	1.00	0.96	0.87 ± 0.04	1.00 ± 0.00	0.93 ± 0.03
Logistic Regression	0.87	1.00	0.94	0.87	1.00	0.93	0.84 ± 0.06	1.00 ± 0.00	0.92 ± 0.04
SVM	0.87	1.00	0.94	0.85	1.00	0.92	0.82 ± 0.07	1.00 ± 0.00	0.91 ± 0.04
Neural Network	1.00	1.00	1.00	0.89	1.00	0.94	0.85 ± 0.05	1.00 ± 0.00	0.90 ± 0.03
Gradient Boosting	1.00	1.00	1.00	0.85	1.00	0.95	0.83 ± 0.05	1.00 ± 0.00	0.91 ± 0.03
Balanced Random Forest	0.93	1.00	0.97	0.88	1.00	0.94	0.86 ± 0.06	1.00 ± 0.00	0.94 ± 0.04

Table 5.2: Classifiers' Precision. Presentation of the Train and Test Precision across all classifiers.

Classifier	Train			Test		
	R	S	RS	R	S	RS
Random Forest	1.00	1.00	1.00	0.9	1.00	0.96
Logistic Regression	0.87	1.00	0.93	0.86	1.00	0.92
SVM	0.87	1.00	0.93	0.84	1.00	0.91
Neural Network	1.00	1.00	1.00	0.89	1.00	0.94
Gradient Boosting	1.00	1.00	1.00	0.85	1.00	0.96
Balanced Random Forest	0.93	1.00	0.97	0.86	1.00	0.92

Table 5.3: Classifiers' Recall. Presentation of the Train and Test Recall across all classifiers.

Classifier	Train			Test		
	R	S	RS	R	S	RS
Random Forest	1.00	1.00	1.00	0.87	1.00	0.96
Logistic Regression	0.9	1.00	0.94	0.89	1.00	0.94
SVM	0.9	1.00	0.94	0.88	1.00	0.94
Neural Network	1.00	1.00	1.00	0.88	1.00	0.93
Gradient Boosting	1.00	1.00	1.00	0.85	1.00	0.95
Balanced Random Forest	0.96	1.00	0.98	0.9	1.00	0.95

Table 5.4: Classifiers' F1-Score. Presentation of the Train and Test F1-score across all classifiers.

Classifier	Train			Test		
	R	S	RS	R	S	RS
Random Forest	1.00	1.00	1.00	0.88	1.00	0.96
Logistic Regression	0.88	1.00	0.94	0.87	1.00	0.93
SVM	0.87	1.00	0.94	0.85	1.00	0.92
Neural Network	1.00	1.00	1.00	0.88	1.00	0.94
Gradient Boosting	1.00	1.00	1.00	0.83	1.00	0.95
Balanced Random Forest	0.94	1.00	0.97	0.88	1.00	0.94

6

Discussion

This thesis aimed to build a generative model able to produce synthetic ECoG data from intra-operative signals derived from severe epileptic patients. To achieve this purpose, the real dataset was clustered into three distinct groups— healthy, injured, and noisy— based on their specific data patterns. Each cluster was then fed into separate generative models, allowing the models to effectively learn and replicate the unique characteristics of each pattern, resulting in the generation of realistic synthetic signals. These signals were evaluated across their time, frequency, and time-frequency domains, successfully capturing the underlying trends of the real data and confirming their reliable fidelity. Furthermore, combining synthetic data with real data significantly improved the classifier's performance, demonstrating the practical utility of the generated signals for medical applications.

6.1 Signal Labelling

During the analysis of the real dataset, it was anticipated that the ioECoG data would exhibit distinct patterns associated with various physiological events, therefore forming different data groups. Indeed, the data revealed three clear clusters, illustrated on the left side of Figure 5.2, each associated with unique signal patterns shown in Figure 5.3. Cluster 1, in particular, was associated to pathological signals due to its specific waveform patterns shown in Figure 5.3. These types of patterns, present throughout the signals in Cluster 1, revealed the presence of epileptiform activity, characterized by distinct spikes and slow waves complexes, similar to those depicted in Figure 2.2 and in the work of Li et al. [12], indicating that this cluster contains signals derived from epileptogenic tissue. Additionally, both this cluster and the cluster labeled as healthy, Cluster 2, align with the data patterns displayed in the study of Gajic et al. [11]. Similarly to what was described in Section 5, Gajic et al. noted that injured signals contained more spaced events with fewer fluctuations, compared to the healthy signals, which exhibited more noise and greater fluctuation. Furthermore, when comparing both plots in Figure 5.2, Cluster 1 and 2 present a strong alignment with the clinician's designated categories of injured and

healthy, reaching similarity percentages of 61.11% and 83.87%, respectively. Moreover, it is important to highlight that, based on the clinicians' classification, Cluster 0 contains an almost equal distribution of healthy and injured signals. The signals contained in this cluster were described as noisy and irregular, forming the 'Noisy' cluster. Actually, this noise could be due to contamination from neighbouring channels within the electrode grid, where different tissue types converge, producing signals that are neither entirely healthy nor entirely pathological - noisy. This phenomenon has been observed in intracranial EEG recordings, where electrodes capture the combined electrical activity of nearby cortical neurons [111], leading to contaminated signals. This contamination could lead, for example, to healthy physiological signals becoming mixed with pathological waveforms [112], resulting in a seemingly noisy signal. In addition, this noise could also be attributed to muscle and machine artifacts, where the first has been referred to having similar features as pathological signals, leading to biased results in the classification of epileptogenic tissue [113]. On top of that, this study was not the first to distinguish between epileptogenic, non-epileptogenic and noisy intracranial EEG data. Nejedly et al. [113] also categorized the data into the same three types of signals: physiological, pathological and artifactual, further reinforcing the validity of the cluster assignment. As such, the alignment with the clinical expertise and the presented conformity with the existing literature strongly supports the validity of the cluster assignments.

Transitioning into the frequency domain of the three clusters, given that patients were under induced deep sleep due to propofol administration during surgery, the frequency analysis was expected to show a higher power in the delta frequency band [114] across all clusters. As anticipated, all clusters exhibited a peak in this band in their mean PSD, as shown in Figure 5.3, though with varying amplitudes depending on the cluster. Additionally, some PSDs presented a peak in the alpha band, which was more common in Cluster 0. This alpha peak can be associated with two factors, both expected to be displayed in ioECoG signals: the use of the propofol itself [115] and the presence of epileptiform discharges [12]. During visual inspection of the signals in section 4.4, those that presented this peak in their PSD were significantly noisier, explaining their assignment to Cluster 0.

6.2 Evaluation of Synthetic Data

6.2.1 Fidelity Analysis

Starting with the evaluation of the signals in their time domain, Figure 5.4 illustrates the comparison of the mean and STD between real and synthetic signals across the three clusters. In all cases, a consistent alignment between the real and synthetic mean signals is observed, with a slightly closer agreement in amplitude and waveform than in the work of Carrle et al. [23], with exception of Cluster 0. This indicates that the synthetic signals in this study closely resemble the real ones in terms of their overall waveform shape, effectively capturing the global temporal patterns of the real data, and therefore demonstrating the

generative model's ability to replicate the amplitude fluctuations and signal trends of the real signals. However, some minor STD deviations are present, although comparable to the results reported by Hartmann et al. [59], they may suggest subtle differences in variability between the synthetic and real signals.

Still in the time domain evaluation, the WD comparisons shown in Figure 5.5 were computed between three data pairings: RR, SS and RS. The RR comparison serves as a baseline, providing reference values for the similarity and variability within the real signals. While the RS comparison directly measures the similarity between real and synthetic signals, the goal is not to achieve the highest similarity between them, but rather to closely resemble the RR values. Having this in consideration, the WD values demonstrated that the synthetic signals maintained the similarity trend within the clusters across all data pairings, with Cluster 1 having the highest WD score, followed by Cluster 2 and then Cluster 0. The higher cost of transforming one signal distribution into another within Cluster 1 is likely due to the distinct epileptiform patterns present throughout the signals, leading to a higher WD. In contrast, the signals within Cluster 0 demonstrated a significant resemblance to one another, yielding lower costs. Although the trends in values were preserved between the RR and SS combinations, the slightly diminished values observed in the latter suggested a reduced variability in the synthetic signals. However, when comparing RS data distributions to the literature, the WD scores surpassed several benchmarks, with values of 0.043 ± 0.025, 0.065 ± 0.045 and 0.045 ± 0.036 for Cluster 0, 1, and 2. These values are notably lower than the WD scores of 0.078 and 0.450 WD reported by Hartmann et al. [59] and Xu et al. [26], respectively. It is important to note, however, that these evaluation metrics compare the RR, RS, and SS distributions, whereas most studies only evaluate the direct RS distributions [59] [24] [57], potentially leading to differences in score interpretation. In addition, further comparison with the literature highlights the greater quality of the generated signals in this study where the generated time-series has a length of 20 seconds, producing more realistic samples compared to the shorter time window's used in most studies - 3 seconds [21], 4 seconds [116], 1 second [37], 0.05 seconds [58] and 8 seconds [23].

In the frequency domain, since most power in ioECoG data was concentrated at lower frequencies, more emphasis was placed on the spike range (0-80Hz) rather than the higher HFO range. Although downsampling the signals to 160 Hz would satisfy the *Nyquist* Theorem [117] while meeting this criteria, and even facilitating the generative process and reducing computational costs, the signals were downsampled to 512 Hz. This value was chosen to retain both spike and ripple ranges (0-250 Hz), ensuring a balanced tradeoff between computational costs and the generation of more realistic samples. Unlike most studies, which downsample to 256Hz [59], 160Hz [22], or 64Hz [37], this study maintained a higher sampling rate, similar to the work of Xiaolong Wu et al. [24], which downsampled to 1000Hz.

Following through the direct comparison of the frequency content, Figure 5.6 showcases this domain between the real and synthetic signals. Due to the higher STD variability

in the real signals within each cluster, the key feature of the synthetic signals is that their power is concentrated in the delta band, similarly to their real versions, rather than having identical amplitude or relative power across slow and fast delta bands. Notably, all synthetic signals in each cluster successfully exhibit a power distribution in the PSD resembling their real counterparts. Specifically, the synthetic Cluster 1 still presents the highest delta power, followed by Cluster 2 and Cluster 0, matching the trend of the real PSD. The results obtained in the frequency domain were consistent to those reported by Park et al. [22] and Hartmann et al. [59], where the synthetic PSD exhibited prominent peaks corresponding to its real counterpart, with slight variations in power. Additionally, Carrle et al. [23] reported synthetic PSDs that did not entirely preserve the prominent peaks found in the real data, possibly revealing the more reliable performance of the generator created in this study. However, the lower STD in the synthetic PSDs also indicates less variability compared with the real signals.

In the evaluation of the time-frequency domain in Figure 5.7 the lower values observed in the RR across all metrics indicate a greater variability in the real data compared to the SS scores, which are the highest. This indicates that the synthetic signals exhibit a higher inter-similarity, meaning that they were unable to replicate the full range of variability present in the real data. This is especially evident in Cluster 0, which showed the worst score in all metrics, but presented the biggest similarity increase when moving from the RR to the SS combination, particularly in SSIM and Pearson's Correlation. This indicates that the random nature of the noisy signals contained in Cluster 0 makes it harder for the generative model to create synthetic signals that fully capture the variability of the real signals. Nevertheless, the differences in the RR and SS pairings are relatively small. For example, the average difference between RR and SS in Cosine Similarity is of only 0.0660 ± 0.0325, and even the biggest variations found in Cluster 0 for SSIM and Pearson's Correlation were only of 0.180 ± 0.051 and 0.275 ± 0.119, respectively. The Mean Squared Error (MSE) difference between RR and SS was also modest at 0.182 ± 0.052. Regarding the RS comparison, Clusters 1 and 2 exhibit moderate-to-high similarity scores across all metrics, such as 0.694 ± 0.087 and 0.672 ± 0.063 for Pearson's Correlation, and 0.757 ± 0.066 and 0.747 ± 0.051 for Cosine Similarity. Although Cluster 0 has the lowest RS scores, they are consistent with the RR pairing. These scores suggest that while the synthetic data does not capture the full variability of the real data, it still maintains a strong correlation with the real signals. In the best of our knowledge, this type of analysis was not found in comparable studies [24], as they computed similarity metrics between signals and not between their time-frequency representations.

6.2.2 Diversity Analysis

The limited variability of the synthetic signals stated until now is confirmed through the Diversity Evaluation results. The overlap between synthetic and real data in the PCA plot indicates that the synthetic signals effectively replicated most of the real patterns.

However, their concentration towards the center of the plot reveals that only part of the diversity was captured, with the tails of the real distributions being slightly overlooked. The congruent t-SNE plots reflect similar findings to the ones reported in the study of Wu et al. [24], although the study reported a greater overlap between data points, pointing out the lack of diversity in the generated samples in this study. This issue can be traced back to a specific cause: during clustering, the data points were grouped based on their characteristics, however some misclassifications occurred, particularly in Cluster 1, resulting in a higher density cluster compared to the others. These misclassifications introduced an unwanted inter-variability in the real data, as reflected in the Diversity results, the STD of the signal means, the PSD means, and even the WD and time-frequency scores. During the generative process, the Critic had to distinguish between real and fake data. That is, when signals deviate significantly from the real data distribution, the Critic classifies them as fake. In turn, misclassifications, having data distributions that differ significantly from the clusters they were assigned to, when presented to the Critic may be consistently labeled as fake. Consequently, the generator learns not to synthesize these types of signals, leading to synthetic data with reduced diversity and variability when compared to the misclassified real data, as demonstrated by the evaluation results.

6.2.3 Utility Analysis

Given the unbalanced nature of the classifiers' dataset, the F1-score is the most appropriate evaluation metric to use [118]. Therefore, it is essential to consider all four metrics, particularly F1-Score, rather than relying solely on Accuracy for a comprehensive evaluation. Actually, all classifiers demonstrated improved performance across the four metrics when using the augmented ioECoG dataset, confirming the utility of the generated signals. Notably, the Gradient Boosting model showed a 12% increase in F1-score, with a minimum improvement of 5% observed across all metrics and classifiers. When compared to other studies in the literature, the average accuracy increase of 6.83% across the six classifiers in this study aligns with the reported ranges: 2.51% [20], 5% [26], 6% [24], with the 10% increase in the Gradient Boosting Classifier matching the range of 79.30% to 96.53% improvement in the CNN classifier from the study of Park et al. [22]. However, the 100% scores across all classifiers trained solely with synthetic data could suggest potential overfitting. This might be explained by the synthetic data's limited diversity, leading to highly distinct clusters. These pronounced differences between clusters likely enhanced the classifiers' ability to distinguish them, thus resulting in significantly higher scores.

6.3 Chapter Conclusions and Future Work

This thesis distinguished the data patterns found in the ioECoG data, a task acknowledged by the literature to be arduous and challenging [15]. The three patterns found,

supported by the research of Nejedly et al. [113], differentiated the signals into epileptogenic and non-epileptogenic, and those contaminated by noise from surrounding tissues or surgical artifacts. The separation of this third group not only enabled the generation of cleaner signals, that were previously corrupted by noise during the generative process, but also facilitated the classification process. Without this separation, the classifier would wrongly classify these signals into either healthy or injured. Now, the ternary classification—healthy, injured, and noisy— enables a more precise distinction between epileptiform activity. In surgical contexts, this classification allows clinicians to review noisy channels and make more informed decisions about resection, providing a more careful and thorough approach to removing injured brain tissue.

Regarding the generation of synthetic ioECoG samples, the DCwGAN-GP model generated strong fidelity signals, resembling the real versions across their time, frequency and time-frequency domain. The improved classification performance with the introduction of synthetic data further validated the utility of these signals, demonstrating their effectiveness in enhancing the classifiers' ability to distinguish epileptogenic tissue.

However, the disparity between the diversity of synthetic and real signals suggests that the clustering analysis may require further refinement to minimize potential misclassifications. This improvement can be achieved by incorporating more discriminative features, such as Wavelet Analysis for spike detection, performing a more rigorous preprocessing of the real signals to effectively remove outliers and artifacts, and using longer time windows to retain more signal information.

Nevertheless, the reduced diversity in the synthetic signals may reflect the noteworthy performance of the DCwGAN-GP model, as the Critic effectively ignored misclassifications in the real data and generated only accurate signals. This highlights the capability of GAN to select signals that represent the majority of the real data distribution. Therefore, as long as most real signals have a reasonable quality, the generator will continue to produce similarly quality signals, while disregarding artifacts or misclassifications. However, it is essential to guarantee that the generated dataset captures all the patterns present in the real data, ensuring that the GAN model is not suffering from mode collapse [119].

Future work built on this thesis can be explored in various directions. First, improving the ground-truth labeling of the signals is essential, not only to enhance the accurate distinction of epileptiform activity but also to facilitate the generative process for synthetic signals. In this regard, Deep Clustering techniques emerge as a promising approach to accurately distinguish between the different patterns and characteristics present in the ioECoG data. Another interesting aspect that could be explored is the temporal relationship and synchronization between signals. Although the signals generated with the DCwGAN-GP model achieved a reliable fidelity compared to the original signals, they did not capture inter-channel dependencies. To address this, techniques such as LSTM networks [57] [60], Recurrent Neural Networks (RNNs) [23] [120], or Transformers [24] could be explored to model temporal dynamics and signal relationships.

7

Conclusion

The aim of this thesis was to develop a generative model capable of synthesizing ioECoG signals from both epileptic and non-epileptic tissue, thereby addressing the challenge of data scarcity in the ioECoG datasets. By generating realistic signals, the synthetic data could augment the available datasets, thus enhancing the robustness of tissue classifiers. To achieve this, a GAN model was proposed: DCwGAN-GP.

Firstly, this study identified three distinct patterns in the ioECoG data using Agglomerative Clustering: healthy, injured, and noisy, where the second demonstrated signs of epileptiform activity.

Next, the proposed DCwGAN-GP was able to generate synthetic data that closely resembled the characteristics of these clusters, effectively replicating the properties of brain tissue patterns.

Following the generation of synthetic ioECoG data, the signals were evaluated for their Fidelity, Diversity, and Utility. In terms of similarity quality measures, the synthetic time-series and PSDs preserved the global waveforms and frequency characteristics of the real data, although lacking in diversity, likely due to unwanted variability in the real dataset.

Finally, the performance comparison of several ML classifiers, trained with and without the augmented dataset, demonstrated consistent improvements when utilizing the synthetic data. This improvement reinforces the utility of generated ioECoG data in enhancing the classifiers' ability to differentiate between healthy, injured, and noisy brain patterns.

These results demonstrate the potential of Generative AI, particularly the proposed DCwGAN-GP model, in generating realistic ioECoG data that can augment existing datasets and improve ML models' ability to distinguish epileptiform activity. By refining the accuracy of these distinctions, this approach may support clinicians in more precisely identifying and resecting the affected brain tissue during surgery, leading to better surgery outcomes and improved quality of life for patients with drug-resistant epilepsy.

Bibliography

[1] J. M. Lourenço. *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. 2021. URL: https://github.com/joaomlourenco/novathesis/raw/main/template.pdf (cit. on p. i).

[2] W. H. Organization. *Epilepsy: a public health imperative*. World Health Organization, 2019, p. 146. URL: https://www.who.int/publications/i/item/epilepsy-a-public-health-imperative (cit. on pp. 1, 3).

[3] A. Consales et al. "The surgical treatment of epilepsy". In: *Neurological Sciences* 42 (6 2021-06), pp. 2249–2260. ISSN: 1590-1874. DOI: 10.1007/s10072-021-05198-y (cit. on pp. 1, 3).

[4] J. S. Duncan and P. N. Taylor. "Optimising epilepsy surgery". In: *The Lancet Neurology* 22 (5 2023-05), pp. 373–374. ISSN: 14744422. DOI: 10.1016/S1474-4422(23)00082-0 (cit. on pp. 1, 3).

[5] T. Yang, S. Hakimian, and T. H. Schwartz. "Intraoperative ElectroCorticoGraphy (ECog): indications, techniques, and utility in epilepsy surgery". In: *Epileptic Disorders* 16 (3 2014-09), pp. 271–279. ISSN: 1294-9361. DOI: 10.1684/epd.2014.0675 (cit. on pp. 1, 3, 4).

[6] I. S. Fernández and T. Loddenkemper. "Electrocorticography for Seizure Foci Mapping in Epilepsy Surgery". In: *Journal of Clinical Neurophysiology* 30.6 (2013), 554–570. DOI: https://doi.org/10.1097/01.wnp.0000436898.10125.70 (cit. on p. 1).

[7] H. M. Greiner et al. "Preresection intraoperative electrocorticography (ECoG) abnormalities predict seizure-onset zone and outcome in pediatric epilepsy surgery". In: *Epilepsia* 57.4 (2016), 582–589. DOI: https://doi.org/10.1111/epi.13341 (cit. on p. 1).

[8] M. Demuru et al. "A Practical Workflow for Organizing Clinical Intraoperative and Long-term iEEG Data in BIDS". In: *Neuroinformatics* 20 (3 2022-07), pp. 727–736. ISSN: 15590089. DOI: 10.1007/s12021-022-09567-6 (cit. on pp. 1, 4).

[9] A. Manole et al. "State of the Art and Challenges in Epilepsy—A Narrative Review". In: *Journal of Personalized Medicine* 13 (4 2023-04), p. 623. ISSN: 2075-4426. DOI: 10.3390/jpm13040623 (cit. on pp. 1, 3).

[10] W. Zweiphenning et al. "Intraoperative electrocorticography using high-frequency oscillations or spikes to tailor epilepsy surgery in the Netherlands (the HFO trial): a randomised, single-blind, adaptive non-inferiority trial". In: *The Lancet Neurology* 21 (11 2022-11), pp. 982–993. ISSN: 14744422. DOI: 10.1016/S1474-4422(22)00311-8 (cit. on pp. 1, 3, 5, 14, 15).

[11] D. Gajic et al. "Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis". In: *Frontiers in Computational Neuroscience* 9 (2015-03). ISSN: 1662-5188. DOI: 10.3389/fncom.2015.00038 (cit. on pp. 1, 4, 8, 9, 34).

[12] Q. Li et al. "Distinguishing Epileptiform Discharges From Normal Electroencephalograms Using Adaptive Fractal and Network Analysis: A Clinical Perspective". In: *Frontiers in Physiology* 11 (2020), p. 828. DOI: 10.3389/fphys.2020.00828. URL: https://doi.org/10.3389/fphys.2020.00828 (cit. on pp. 1, 5, 34, 35).

[13] S. Saminu et al. "Application of Deep Learning and WT-SST in Localization of Epileptogenic Zone Using Epileptic EEG Signals". In: *Applied Sciences* 12 (10 2022-05), p. 4879. ISSN: 2076-3417. DOI: 10.3390/app12104879 (cit. on pp. 1, 8, 9).

[14] M. Shen et al. "Epileptic Tissue Localization through Skewness-Based Functional Connectivity in the High-Frequency Band of Intracranial EEG". In: *Bioengineering* 10 (4 2023-04), p. 461. ISSN: 2306-5354. DOI: 10.3390/bioengineering10040461 (cit. on pp. 1, 5, 8, 9).

[15] S. Chaibi et al. "Epileptic EEG patterns recognition through machine learning techniques and relevant time–frequency features". In: *Biomedical Engineering / Biomedizinische Technik* 69 (2 2024-04), pp. 111–123. ISSN: 0013-5585. DOI: 10.1515/bmt-2023-0332 (cit. on pp. 1, 8, 9, 15, 16, 38).

[16] C. Silva Lourenço, M. C. Tjepkema-Cloostermans, and M. J.A.M. "Efficient use of clinical EEG data for deep learning in epilepsy". In: *Clinical Neurophysiology* 132.6 (2021), 1234–1240. DOI: https://doi.org/10.1016/j.clinph.2021.01.035 (cit. on p. 1).

[17] F. Escobar-Ipuz et al. "Prediction of patients with idiopathic generalized epilepsy from healthy controls using machine learning from scalp EEG recordings". In: *Brain Research* 1798 (2023), p. 148131. DOI: https://doi.org/10.1016/j.brainres.2022.148131 (cit. on p. 1).

[18] S. Liang et al. "An Auxiliary Synthesis Framework for Enhancing EEG-Based Classification With Limited Data". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31 (2023), pp. 2120–2131. ISSN: 1534-4320. DOI: 10.1109/TNSRE.2023.3268979 (cit. on pp. 1, 9–11).

[19] K. Shu et al. *Data Augmentation for Seizure Prediction with Generative Diffusion Model*. 2023. arXiv: 2306.08256 [eess.SP]. URL: https://arxiv.org/abs/2306.08256 (cit. on pp. 1, 5, 7).

[20] X. Du et al. "A Data Augmentation Method for Motor Imagery EEG Signals Based on DCGAN-GP Network". In: *Brain Sciences* 14 (4 2024-04), p. 375. ISSN: 2076-3425. DOI: 10.3390/brainsci14040375 (cit. on pp. 2, 8, 11, 38).

[21] N. K. N. Aznan et al. "Simulating Brain Signals: Creating Synthetic EEG Data via Neural-Based Generative Models for Improved SSVEP Classification". In: (2019-01). DOI: 10.1109/IJCNN.2019.8852227. URL: http://arxiv.org/abs/1901.07429http://dx.doi.org/10.1109/IJCNN.2019.8852227 (cit. on pp. 2, 8, 9, 18, 19, 36).

[22] J. Park, P. Mahey, and O. Adeniyi. *Improving EEG Signal Classification Accuracy Using Wasserstein Generative Adversarial Networks*. 2024. arXiv: 2402.09453 [eess.SP]. URL: https://arxiv.org/abs/2402.09453 (cit. on pp. 2, 5, 6, 8, 9, 19, 36–38).

[23] F. P. Carrle, Y. Hollenbenders, and A. Reichenbach. "Generation of synthetic EEG data for training algorithms supporting the diagnosis of major depressive disorder". In: *Frontiers in Neuroscience* 17 (2023). ISSN: 1662453X. DOI: 10.3389/fnins.2023.1219133 (cit. on pp. 2, 7, 9, 19, 35–37, 39).

[24] X. Wu et al. "Data augmentation for invasive brain-computer interfaces based on stereo-electroencephalography (SEEG)". In: *Journal of Neural Engineering* (2024-01). ISSN: 1741-2560. DOI: 10.1088/1741-2552/ad200e (cit. on pp. 2, 6, 9–11, 13, 22, 36–39).

[25] A. A. E. Shoka et al. "EEG seizure detection: concepts, techniques, challenges, and future trends". In: *Multimedia Tools and Applications* 82 (27 2023-11), pp. 42021–42051. ISSN: 1380-7501. DOI: 10.1007/s11042-023-15052-2 (cit. on p. 3).

[26] Y. Xu, J. Yang, and M. Sawan. "Multichannel Synthetic Preictal EEG Signals to Enhance the Prediction of Epileptic Seizures". In: *IEEE Transactions on Biomedical Engineering* 69 (11 2022-11), pp. 3516–3525. ISSN: 0018-9294. DOI: 10.1109/TBME.2022.3171982 (cit. on pp. 3, 10, 36, 38).

[27] NHS. *Treatment - Epilepsy*. NHS, 2020-09. URL: https://www.nhs.uk/conditions/epilepsy/treatment/ (cit. on p. 3).

[28] J. Lopez et al. "Drug-resistant epilepsy: Definition and treatment alternatives". In: *Neurologia (Barcelona, Spain)* 30 (2014-06). DOI: 10.1016/j.nrleng.2014.04.002 (cit. on p. 3).

[29] W. O. Tatum, ed. *Handbook of EEG Interpretation*. Springer Publishing Company, 2021-05. ISBN: 978-0-8261-4708-0. DOI: 10.1891/9780826147097 (cit. on pp. 4, 14).

[30] M. R. Mercier et al. "Advances in human intracranial electroencephalography research, guidelines and good practices". In: *NeuroImage* 260 (2022), p. 119438. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2022.119438. URL: https://www.sciencedirect.com/science/article/pii/S1053811922005559 (cit. on p. 4).

[31] M. E. C. Andraus and S. V. Alves-Leon. "Non-epileptiform EEG abnormalities: an overview". In: *Arquivos de Neuro-Psiquiatria* 69 (5 2011-10), pp. 829–835. ISSN: 0004-282X. DOI: 10.1590/S0004-282X2011000600020 (cit. on p. 4).

[32] W. J. E. M. Zweiphenning et al. "Increased gamma and decreased fast ripple connections of epileptic tissue: A high-frequency directed network approach". In: *Epilepsia* 60.9 (2019), 1908–1920. DOI: https://doi.org/10.1111/epi.16296 (cit. on pp. 5, 15).

[33] IBM. *What Is Deep Learning?* URL: https://www.ibm.com/topics/deep-learning (cit. on p. 5).

[34] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning". In: *Nature* 521 (7553 2015-05), pp. 436–444. ISSN: 0028-0836. DOI: 10.1038/nature14539 (cit. on p. 5).

[35] K. Perkonoja, K. Auranen, and J. Virta. *Methods for generating and evaluating synthetic longitudinal patient data: a systematic review*. 2024. arXiv: 2309.12380 [stat.ME]. URL: https://arxiv.org/abs/2309.12380 (cit. on pp. 5, 6).

[36] A. Goncalves et al. "Generation and evaluation of synthetic patient data". In: *BMC Medical Research Methodology* 20 (1 2020-12), p. 108. ISSN: 1471-2288. DOI: 10.1186/s12874-020-00977-1 (cit. on p. 5).

[37] S. Panwar et al. "Modeling EEG Data Distribution With a Wasserstein Generative Adversarial Network to Predict RSVP Events". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28 (8 2020-08), pp. 1720–1730. ISSN: 1534-4320. DOI: 10.1109/TNSRE.2020.3006180 (cit. on pp. 5, 6, 8, 9, 11, 36).

[38] I. Goodfellow. *NIPS 2016 Tutorial: Generative Adversarial Networks*. 2017. arXiv: 1701.00160 [cs.LG]. URL: https://arxiv.org/abs/1701.00160 (cit. on p. 5).

[39] N. K. N. Aznan et al. "Leveraging Synthetic Subject Invariant EEG Signals for Zero Calibration BCI". In: (2020-07). URL: http://arxiv.org/abs/2007.11544 (cit. on pp. 6, 8, 10).

[40] I. J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML]. URL: https://arxiv.org/abs/1406.2661 (cit. on p. 6).

[41] X. Jin et al. "Chapter 2 - Game Theory for Infrastructure Security: The Power of Intent-Based Adversary Models11Part of the results included in this chapter was presented in [1], [2]." In: *Handbook on Securing Cyber-Physical Critical Infrastructure*. Ed. by S. K. Das, K. Kant, and N. Zhang. Boston: Morgan Kaufmann, 2012, pp. 31–53. ISBN: 978-0-12-415815-3. DOI: https://doi.org/10.1016/B978-0-12-415815

-3.00002-9. URL: https://www.sciencedirect.com/science/article/pii/B9780124158153000029 (cit. on p. 6).

[42] D. Pascual et al. "EpilepsyGAN: Synthetic Epileptic Brain Activities With Privacy Preservation". In: *IEEE Transactions on Biomedical Engineering* 68 (8 2021-08), pp. 2435–2446. ISSN: 0018-9294. DOI: 10.1109/TBME.2020.3042574 (cit. on pp. 6, 9).

[43] L. Yang et al. "Diffusion Models: A Comprehensive Survey of Methods and Applications". In: *ACM Computing Surveys* 56 (4 2024-04), pp. 1–39. ISSN: 0360-0300. DOI: 10.1145/3626235 (cit. on p. 6).

[44] A. Figueira and B. Vaz. *Survey on Synthetic Data Generation, Evaluation Methods and GANs*. 2022-08. DOI: 10.3390/math10152733 (cit. on p. 7).

[45] A. M. Alaa et al. *How Faithful is your Synthetic Data? Sample-level Metrics for Evaluating and Auditing Generative Models*. 2022. arXiv: 2102.08921 [cs.LG]. URL: https://arxiv.org/abs/2102.08921 (cit. on p. 7).

[46] M. F. Naeem et al. *Reliable Fidelity and Diversity Metrics for Generative Models*. 2020. arXiv: 2002.09797 [cs.CV]. URL: https://arxiv.org/abs/2002.09797 (cit. on p. 7).

[47] M. Demuru et al. "The value of intra-operative electrographic biomarkers for tailoring during epilepsy surgery: from group-level to patient-level analysis". In: *Scientific Reports* 10.1 (2020-09), p. 14654. ISSN: 2045-2322. DOI: 10.1038/s41598-020-71359-2. URL: https://doi.org/10.1038/s41598-020-71359-2 (cit. on p. 8).

[48] D. Pascual et al. "Synthetic Epileptic Brain Activities Using Generative Adversarial Networks". In: (2019-07). URL: http://arxiv.org/abs/1907.10518 (cit. on pp. 8, 10).

[49] M. Hu et al. "E2SGAN: EEG-to-SEEG translation with generative adversarial networks". In: *Frontiers in Neuroscience* 16 (2022-09). ISSN: 1662-453X. DOI: 10.3389/fnins.2022.971829 (cit. on p. 8).

[50] L. Fang et al. "Enhancing Medical Signal Processing and Diagnosis with AI-Generated Content Techniques". In: *IEEE Journal of Biomedical and Health Informatics* (2024). ISSN: 21682208. DOI: 10.1109/JBHI.2024.3429560 (cit. on pp. 9, 11, 24).

[51] A. F. Nia et al. "Synthesizing affective neurophysiological signals using generative models: A review paper". In: *Journal of Neuroscience Methods* 406 (2024-06). ISSN: 1872678X. DOI: 10.1016/j.jneumeth.2024.110129 (cit. on p. 9).

[52] A. G. Habashi et al. "Generative adversarial networks in EEG analysis: an overview". In: *Journal of NeuroEngineering and Rehabilitation* 20 (1 2023-04), p. 40. ISSN: 1743-0003. DOI: 10.1186/s12984-023-01169-w (cit. on p. 9).

[53] A. Science and C. S. Publications. *A Dual-Discriminator GAN for Sleep EEG Signal Synthesis*. Version Volume 10, Issue 1. 2024-01. DOI: 10.5281/zenodo.10553550. URL: https://doi.org/10.5281/zenodo.10553550 (cit. on pp. 9–11).

[54] M. Mirza and S. Osindero. *Conditional Generative Adversarial Nets*. 2014. arXiv: 1411.1784 [cs.LG]. URL: https://arxiv.org/abs/1411.1784 (cit. on p. 9).

[55] K. Rasheed et al. "A Generative Model to Synthesize EEG Data for Epileptic Seizure Prediction". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), pp. 2322–2332. ISSN: 15580210. DOI: 10.1109/TNSRE.2021.3125023 (cit. on pp. 9, 11).

[56] K. Lalli and M. Senbagavalli. "Enhancing Deep Learning for Autism Spectrum Disorder Detection with Dual-Encoder GAN-based Augmentation of Electroencephalogram Data". In: *Salud, Ciencia y Tecnologia - Serie de Conferencias* 3 (2024-03). ISSN: 29534860. DOI: 10.56294/sctconf2024958 (cit. on p. 10).

[57] X. Du et al. "Electroencephalographic Signal Data Augmentation Based on Improved Generative Adversarial Network". In: *Brain Sciences* 14 (4 2024-04). ISSN: 20763425. DOI: 10.3390/brainsci14040367 (cit. on pp. 10, 11, 36, 39).

[58] Y. Pan et al. "Short-length SSVEP data extension by a novel generative adversarial networks based framework". In: (2023-01). URL: http://arxiv.org/abs/2301.05599 (cit. on pp. 10, 36).

[59] K. G. Hartmann, R. T. Schirrmeister, and T. Ball. "EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals". In: (2018-06). URL: http://arxiv.org/abs/1806.01875 (cit. on pp. 10, 11, 36, 37).

[60] Z. Cook et al. "Enhancing Brain Age Prediction: A Generative AI Approach for EEG Machine Learning Models". In: *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2024-05, pp. 1–6. ISBN: 979-8-3503-8090-3. DOI: 10.1109/I2MTC60896.2024.10560863 (cit. on pp. 11, 39).

[61] A. Delorme. "EEG is better left alone". In: *Scientific Reports* 13 (1 2023-02), p. 2372. ISSN: 2045-2322. DOI: 10.1038/s41598-023-27528-0 (cit. on p. 13).

[63] S. S. Rasheed and F. S. Miften. "Improve of Neonatal Seizure Detection based on EEG Signal using K-mean clustering". In: *2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT)*. IEEE, 2023-07, pp. 181–184. ISBN: 979-8-3503-4188-1. DOI: 10.1109/AICCIT57614.2023.10218158 (cit. on pp. 13, 15).

[64] K. J. Miller. "A library of human electrocorticographic data and analyses". In: *Nature Human Behaviour* 3 (11 2019-08), pp. 1225–1235. ISSN: 2397-3374. DOI: 10.1038/s41562-019-0678-3 (cit. on p. 14).

[62] N. van Klink et al. "High frequency oscillations in intra-operative electrocorticography before and after epilepsy surgery". In: *Clinical Neurophysiology* 125.11 (2014), pp. 2212–2219. ISSN: 1388-2457. DOI: https://doi.org/10.1016/j.clinph.2014.03.004. URL: https://www.sciencedirect.com/science/article/pii/S1388245714001291 (cit. on pp. 14, 15).

[65] T. S. community. *SciPy API - Signal processing (scipy.signal) - iirnotch*. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.iirnotch.html (cit. on p. 14).

[66] H. Zhang et al. "The applied principles of EEG analysis methods in neuroscience and clinical neurology". In: *Military Medical Research* 10 (1 2023-12), p. 67. ISSN: 2054-9369. DOI: 10.1186/s40779-023-00502-7 (cit. on p. 14).

[67] S. A. Unde and R. Shriram. "Coherence Analysis of EEG Signal Using Power Spectral Density". In: *2014 Fourth International Conference on Communication Systems and Network Technologies*. 2014, pp. 871–874. DOI: 10.1109/CSNT.2014.181 (cit. on p. 14).

[68] H. W. Shin et al. "Monitoring of anesthetic depth and EEG band power using phase lag entropy during propofol anesthesia". In: *BMC Anesthesiology* 20 (1 2020-12), p. 49. ISSN: 1471-2253. DOI: 10.1186/s12871-020-00964-5 (cit. on pp. 14, 16).

[69] S. Wickramaratne, A. Parekh, and K. Kam. "0319 Using Generative AI as a Tool for Simulating EEG During Nocturnal Polysomnography". In: *SLEEP* 47 (Supplement$_1$ 2024-04), A137–A137. ISSN: 0161-8105. DOI: 10.1093/sleep/zsae067.0319 (cit. on p. 14).

[70] P. Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2 (cit. on pp. 14, 15, 22).

[71] SciPy Community. *Signal Processing (scipy.signal)*. 2024. URL: https://docs.scipy.org/doc/scipy/tutorial/signal.html#spectral-analysis (cit. on p. 14).

[72] MATLAB Simulink. *Power Spectral Density Estimates Using FFT*. https://www.mathworks.com/help/signal/ug/power-spectral-density-estimates-using-fft.html (cit. on p. 14).

[73] M. Barandas et al. "TSFEL: Time Series Feature Extraction Library". In: *SoftwareX* 11 (2020), p. 100456. ISSN: 2352-7110. DOI: https://doi.org/10.1016/j.softx.2020.100456. URL: https://www.sciencedirect.com/science/article/pii/S2352711020300017 (cit. on pp. 15, 16).

[74] C. R. Harris et al. "Array programming with NumPy". In: *Nature* 585.7825 (2020-09), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2 (cit. on p. 15).

[75] Y. Du et al. "Unsupervised Multivariate Feature-Based Adaptive Clustering Analysis of Epileptic EEG Signals". In: *Brain Sciences* 14 (4 2024-03), p. 342. ISSN: 2076-3425. DOI: 10.3390/brainsci14040342 (cit. on p. 15).

[76] D. O'Shaughnessy. "Review of analysis methods for speech applications". In: *Speech Communication* 151 (2023), pp. 64–75. ISSN: 0167-6393. DOI: https://doi.org/10.1016/j.specom.2023.05.008. URL: https://www.sciencedirect.com/science/article/pii/S0167639323000766 (cit. on p. 16).

[77] G. Espinosa-Paredes. "Chapter 7 - Nonlinear BWR dynamics with a fractional reduced order model". In: *Fractional-Order Models for Nuclear Reactor Analysis*. Ed. by G. Espinosa-Paredes. Woodhead Publishing Series in Energy. Woodhead Publishing, 2021, pp. 247–295. ISBN: 978-0-12-823665-9. DOI: https://doi.org/10.1016/B978-0-12-823665-9.00007-9. URL: https://www.sciencedirect.com/science/article/pii/B9780128236659000079 (cit. on p. 16).

[78] H. A. Díaz M. and F. Córdova. "On the meaning of Hurst entropy applied to EEG data series". In: *Procedia Computer Science* 199 (2022). The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 2021): Developing Global Digital Economy after COVID-19, pp. 1385–1392. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2022.01.175. URL: https://www.sciencedirect.com/science/article/pii/S1877050922001764 (cit. on p. 16).

[79] J. Semmlow. "Chapter 2 - Basic Concepts in Signal Processing". In: *Signals and Systems for Bioengineers (Second Edition)*. Ed. by J. Semmlow. Second Edition. Biomedical Engineering. Boston: Academic Press, 2012, pp. 35–80. ISBN: 978-0-12-384982-3. DOI: https://doi.org/10.1016/B978-0-12-384982-3.00002-X. URL: https://www.sciencedirect.com/science/article/pii/B978012384982300002X (cit. on p. 16).

[80] F. Pedregosa et al. *Scikit-learn: Machine Learning in Python*. 2018. arXiv: 1201.0490 [cs.LG]. URL: https://arxiv.org/abs/1201.0490 (cit. on pp. 15–17, 23, 24).

[81] S.-L. D. B. License). *Standardscaler*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html (cit. on p. 15).

[82] M. C. Guerrero, J. S. Parada, and H. E. Espitia. "EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks". In: *Heliyon* 7 (6 2021-06), e07258. ISSN: 24058440. DOI: 10.1016/j.heliyon.2021.e07258 (cit. on p. 15).

[83] T. D.K., P. B.G, and F. Xiong. "Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques". In: *Pattern Recognition Letters* 128 (2019), pp. 544–550. ISSN: 0167-8655. DOI: https://doi.org/10.10

16/j.patrec.2019.10.029. URL: https://www.sciencedirect.com/science/article/pii/S0167865519303083 (cit. on p. 15).

[84] H. Zhou, X. Wang, and R. Zhu. "Feature selection based on mutual information with correlation coefficient". In: *Applied Intelligence* 52 (5 2022-03), pp. 5457–5474. ISSN: 0924-669X. DOI: 10.1007/s10489-021-02524-x (cit. on p. 15).

[85] A. J. Onumanyi et al. "AutoElbow: An Automatic Elbow Detection Method for Estimating the Number of Clusters in a Dataset". In: *Applied Sciences* 12.15 (2022). ISSN: 2076-3417. DOI: 10.3390/app12157515. URL: https://www.mdpi.com/2076-3417/12/15/7515 (cit. on p. 16).

[86] scikit-learn Community. *Silhouette*$_{S}$*core*. 2024. URL: https://scikit-learn.org/1.5/modules/generated/sklearn.metrics.silhouette_score.html (cit. on p. 17).

[87] K. R. Shahapure and C. Nicholas. "Cluster Quality Analysis Using Silhouette Score". In: *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2020-10, pp. 747–748. ISBN: 978-1-7281-8206-3. DOI: 10.1109/DSAA49011.2020.00096 (cit. on p. 17).

[88] I. Kononenko and M. Kukar. "Chapter 12 - Cluster Analysis". In: *Machine Learning and Data Mining*. Ed. by I. Kononenko and M. Kukar. Woodhead Publishing, 2007, pp. 321–358. ISBN: 978-1-904275-21-3. DOI: https://doi.org/10.1533/9780857099440.321. URL: https://www.sciencedirect.com/science/article/pii/B9781904275213500125 (cit. on p. 17).

[89] A. Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf (cit. on p. 17).

[90] NVIDIA Corporation. *NVIDIA RTX 6000 Ada Generation Graphics Card*. https://www.nvidia.com/en-us/design-visualization/rtx-6000/. 2024 (cit. on p. 17).

[91] H. Zhong et al. "Fine-tuning transfer learning based on DCGAN integrated with self-attention and spectral normalization for bearing fault diagnosis". In: *Measurement* 210 (2023), p. 112421. ISSN: 0263-2241. DOI: https://doi.org/10.1016/j.measurement.2022.112421. URL: https://www.sciencedirect.com/science/article/pii/S0263224122016189 (cit. on p. 19).

[92] *Zeroing out gradients in PyTorch*. https://pytorch.org/tutorials/recipes/recipes/zeroing_out_gradients.html. Zeroing out gradients in PyTorch - PyTorch Tutorials 2.4.0+cu121 documentation. 2024 (cit. on p. 19).

[93] SciPy Community. *Wasserstein_Distance*. 2024. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wasserstein_distance.html (cit. on p. 21).

[94] MathWorks. *Wavelet Time-Frequency Analyzer*. 2024. URL: https://www.mathworks.com/help/wavelet/ug/time-frequency-analysis-and-continuous-wavelet-transform.html (cit. on p. 22).

[95] MathWorks. *Wavelet Transforms in MATLAB*. 2024. URL: https://www.mathworks.com/discovery/wavelet-transforms.html (cit. on p. 22).

[96] J. T. Bialasiewicz. "Application of Wavelet Scalogram and Coscalogram for Analysis of Biomedical Signals". In: 2015. URL: https://api.semanticscholar.org/CorpusID:11995759 (cit. on p. 22).

[97] M. M. Mukaka. "Statistics corner: A guide to appropriate use of correlation coefficient in medical research". In: *Malawi medical journal: the journal of Medical Association of Malawi* 24.3 (2012), pp. 69–71. URL: https://pubmed.ncbi.nlm.nih.gov/23638278/ (cit. on p. 22).

[98] J. Han, M. Kamber, and J. Pei. "2 - Getting to Know Your Data". In: *Data Mining (Third Edition)*. Ed. by J. Han, M. Kamber, and J. Pei. Third Edition. The Morgan Kaufmann Series in Data Management Systems. Boston: Morgan Kaufmann, 2012, pp. 39–82. ISBN: 978-0-12-381479-1. DOI: https://doi.org/10.1016/B978-0-12-381479-1.00002-2. URL: https://www.sciencedirect.com/science/article/pii/B9780123814791000022 (cit. on p. 22).

[99] B. Li and L. Han. "Distance Weighted Cosine Similarity Measure for Text Classification". In: 2013, pp. 611–618. DOI: 10.1007/978-3-642-41278-3_74 (cit. on p. 22).

[100] I. Bakurov et al. "Structural similarity index (SSIM) revisited: A data-driven approach". In: *Expert Systems with Applications* 189 (2022), p. 116087. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2021.116087. URL: https://www.sciencedirect.com/science/article/pii/S0957417421014238 (cit. on p. 22).

[101] S. van der Walt et al. "scikit-image: image processing in Python". In: *PeerJ* 2 (2014-06), e453. ISSN: 2167-8359. DOI: 10.7717/peerj.453. URL: http://dx.doi.org/10.7717/peerj.453 (cit. on p. 22).

[102] G. Harish Babu and N. Venkatram. "A survey on analysis and implementation of state-of-the-art haze removal techniques". In: *Journal of Visual Communication and Image Representation* 72 (2020), p. 102912. ISSN: 1047-3203. DOI: https://doi.org/10.1016/j.jvcir.2020.102912. URL: https://www.sciencedirect.com/science/article/pii/S1047320320301504 (cit. on p. 22).

[103] A. A. Awan. *Introduction to T-SNE: Nonlinear dimensionality reduction and Data Visualization*. DataCamp, https://www.datacamp.com/tutorial/introduction-t-sne. 2023. URL: https://www.datacamp.com/tutorial/introduction-t-sne (cit. on pp. 22, 23).

[104] I. T. Jolliffe and J. Cadima. "Principal component analysis: a review and recent developments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2065 2016-04), p. 20150202. ISSN: 1364-503X. DOI: 10.1098/rsta.2015.0202 (cit. on p. 23).

[105] H. A. Salman, A. Kalakech, and A. Steiti. "Random Forest Algorithm Overview". In: *Babylonian Journal of Machine Learning* 2024 (2024-06), pp. 69–79. ISSN: 3006-5429. DOI: 10.58496/BJML/2024/007 (cit. on p. 23).

[106] A. Géron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. "O'Reilly Media, Inc.", 2017-03. URL: https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/#:~:text=This%20practical%20book%20shows%20you%20how.%20By%20using%20concrete%20examples, (cit. on pp. 23, 24).

[107] M. M. Adankon and M. Cheriet. "Support Vector Machine". In: *Encyclopedia of Biometrics*. Ed. by S. Z. Li and A. Jain. Boston, MA: Springer US, 2009, pp. 1303–1308. ISBN: 978-0-387-73003-5. DOI: 10.1007/978-0-387-73003-5_299. URL: https://doi.org/10.1007/978-0-387-73003-5_299 (cit. on p. 23).

[108] S. Jaiswal. *Multilayer Perceptrons in Machine Learning: A Comprehensive Guide*. 2024. URL: https://www.datacamp.com/tutorial/multilayer-perceptrons-in-machine-learning (cit. on p. 23).

[109] B. Tuychiev. *A Guide to the Gradient Boosting Algorithm*. https://www.datacamp.com/tutorial/guide-to-the-gradient-boosting-algorithm. Accessed: 2024-09-07. 2023. URL: https://www.datacamp.com/tutorial/guide-to-the-gradient-boosting-algorithm (cit. on p. 23).

[110] C. Chen. "Using Random Forest to Learn Imbalanced Data". In: 2004. URL: https://api.semanticscholar.org/CorpusID:7308660 (cit. on p. 23).

[111] K. R. Sindhu et al. "A novel method for dynamically altering the surface area of intracranial EEG electrodes". In: *Journal of Neural Engineering* 20.2 (2023), p. 026002. DOI: 10.1088/1741-2552/acb79f (cit. on p. 35).

[112] E. F. Chang. "Data quality, epilepsy, interictal discharges, intracranial recordings, and signal processing". In: *Frontiers in Human Neuroscience* 14 (2020), p. 44. DOI: 10.3389/fnhum.2020.00044. URL: https://doi.org/10.3389/fnhum.2020.00044 (cit. on p. 35).

[113] P. Nejedly et al. "Multicenter intracranial EEG dataset for classification of graphoelements and artifactual signals". In: *Scientific Data* 7.1 (2020), p. 179. ISSN: 2052-4463. DOI: 10.1038/s41597-020-0532-5. URL: https://doi.org/10.1038/s41597-020-0532-5 (cit. on pp. 35, 39).

[114] J. Moini and P. Piran. "Chapter 6 - Cerebral cortex". In: *Functional and Clinical Neuroanatomy*. Ed. by J. Moini and P. Piran. Academic Press, 2020, pp. 177–240. ISBN: 978-0-12-817424-1. DOI: https://doi.org/10.1016/B978-0-12-817424-1.00006-9. URL: https://www.sciencedirect.com/science/article/pii/B9780128174241000069 (cit. on p. 35).

[115] P. L. Purdon et al. "Clinical Electroencephalography for Anesthesiologists: Part I: Background and Basic Signatures". In: *Anesthesiology* 123.4 (2015-10), pp. 937–960. ISSN: 0003-3022. DOI: 10.1097/ALN.0000000000000841. eprint: https://pubs.asahq.org/anesthesiology/article-pdf/123/4/937/373943/20151000_0-00034.pdf. URL: https://doi.org/10.1097/ALN.0000000000000841 (cit. on p. 35).

[116] D. Pascual et al. *Synthetic Epileptic Brain Activities Using Generative Adversarial Networks*. 2019. arXiv: 1907.10518 [eess.SP]. URL: https://arxiv.org/abs/1907.10518 (cit. on p. 36).

[117] R. Oshana. "4 - Overview of Digital Signal Processing Algorithms". In: *DSP Software Development Techniques for Embedded and Real-Time Systems*. Ed. by R. Oshana. Embedded Technology. Burlington: Newnes, 2006, pp. 59–121. ISBN: 978-0-7506-7759-2. DOI: https://doi.org/10.1016/B978-075067759-2/50006-5. URL: https://www.sciencedirect.com/science/article/pii/B9780750677592500065 (cit. on p. 36).

[118] S. Riyanto et al. "Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification". In: *International Journal of Advanced Computer Science and Applications* 14.6 (2023). DOI: 10.14569/IJACSA.2023.01406116. URL: http://dx.doi.org/10.14569/IJACSA.2023.01406116 (cit. on p. 38).

[119] A. Murray and D. B. Rawat. "On the Performance of Generative Adversarial Network by Limiting Mode Collapse for Malware Detection Systems". In: *Sensors (Basel, Switzerland)* 22.1 (2021), p. 264. DOI: 10.3390/s22010264. URL: https://doi.org/10.3390/s22010264 (cit. on p. 39).

[120] D. Zhang. "A Comprehensive Study on Robust EEG Signal Generation and Evaluation". In: *The 2nd International Conference on Computing and Data Science*. CONF-CDS 2021. Stanford, CA, USA: Association for Computing Machinery, 2021. ISBN: 9781450389570. DOI: 10.1145/3448734.3450467. URL: https://doi.org/10.1145/3448734.3450467 (cit. on p. 39).

I

Loss Functions

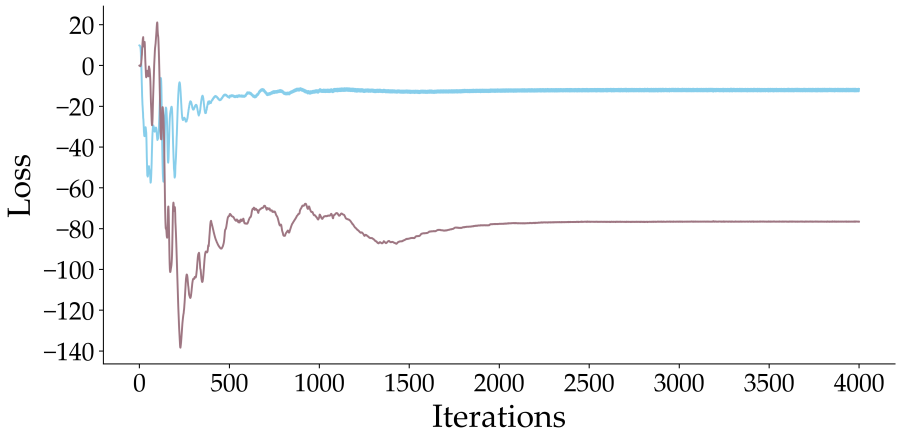

Figure I.1: Loss Function of the Generative Model for Cluster 0. The blue line represents the variation in the Critic's loss function, while the purple line corresponds to the Generator's loss function.

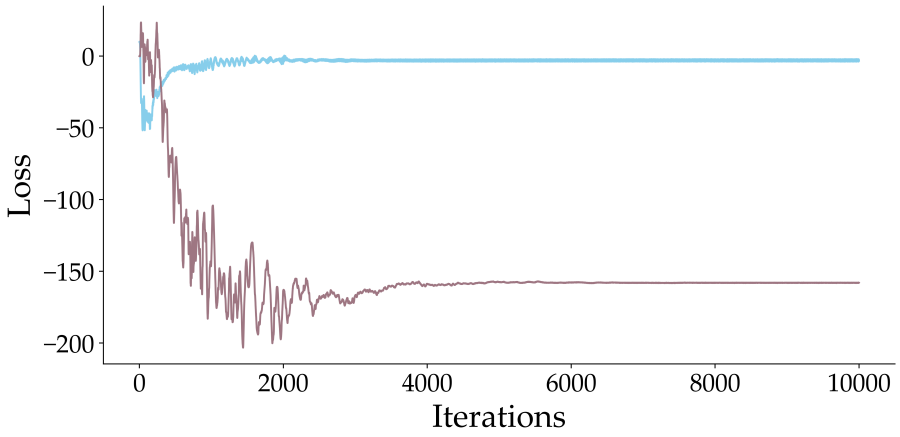

Figure I.2: Loss Function of the Generative Model for Cluster 1. The blue line represents the variation in the Critic's loss function, while the purple line corresponds to the Generator's loss function.

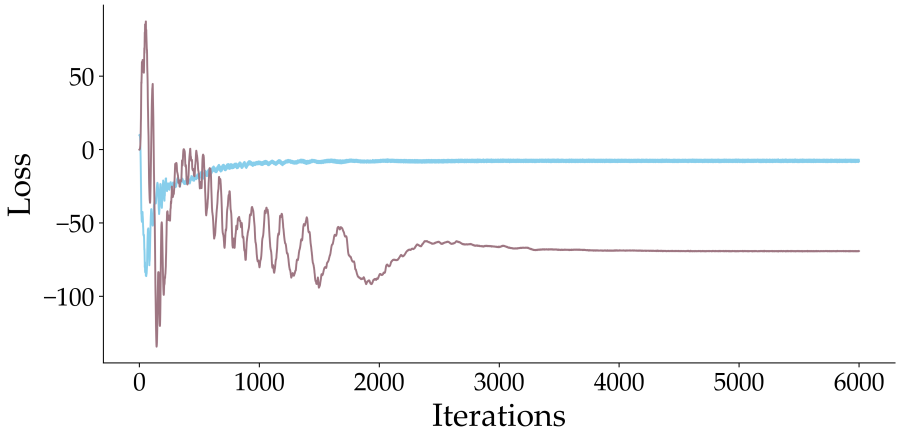

Figure I.3: Loss Function of the Generative Model for Cluster 2. The blue line represents the variation in the Critic's loss function, while the purple line corresponds to the Generator's loss function.